Graduate Texts in Mathematics

Graduate Texts in Mathematics

Serge Lang

Algebra

Revised Third Edition

Springer

Graduate Texts in Mathematics 211

Springer
New York
Berlin
Heidelberg
Barcelona
Hong Kong
London
Milan
Paris
Singapore
Tokyo

Serge Lang
Department of Mathematics
Yale University
New Haven, CT 96520
USA

Mathematics Subject Classification (2000): 13-01, 15-01, 16-01, 20-01

Library of Congress Cataloging-in-Publication Data
Algebra / Serge Lang.—Rev. 3rd ed.
p. cm. — (Graduate texts in mathematics ; 211)
Includes bibliographical references and index.
ISBN 0-387-95385-X (alk. paper)
1. Algebra. I. Title. II. Series.
QA154.3.L3 2002
512—dc21 2001054916

Printed on acid-free paper.

This title was previously published by Addison-Wesley, Reading, MA 1993.

Production managed by Terry Kornak; manufacturing supervised by Erica Bresler.
Revisions typeset by Asco Typesetters, North Point, Hong Kong.
Printed and bound by Edwards Brothers, Inc., Ann Arbor, MI.
Printed in the United States of America.

9 8 7 6 5 4 3 2 1

ISBN 0-387-95385-X SPIN 10855619

Springer-Verlag New York Berlin Heidelberg
A member of BertelsmannSpringer Science+Business Media GmbH

FOREWORD

The present book is meant as a basic text for a one-year course in algebra, at the graduate level.

A perspective on algebra

As I see it, the graduate course in algebra must primarily prepare students to handle the algebra which they will meet in all of mathematics: topology, partial differential equations, differential geometry, algebraic geometry, analysis, and representation theory, not to speak of algebra itself and algebraic number theory with all its ramifications. Hence I have inserted throughout references to papers and books which have appeared during the last decades, to indicate some of the directions in which the algebraic foundations provided by this book are used; I have accompanied these references with some motivating comments, to explain how the topics of the present book fit into the mathematics that is to come subsequently in various fields; and I have also mentioned some unsolved problems of mathematics in algebra and number theory. The *abc* conjecture is perhaps the most spectacular of these.

Often when such comments and examples occur out of the logical order, especially with examples from other branches of mathematics, of necessity some terms may not be defined, or may be defined only later in the book. I have tried to help the reader not only by making cross-references within the book, but also by referring to other books or papers which I mention explicitly.

I have also added a number of exercises. On the whole, I have tried to make the exercises complement the examples, and to give them aesthetic appeal. I have tried to use the exercises also to drive readers toward variations and applications of the main text, as well as toward working out special cases, and as openings toward applications beyond this book.

Organization

Unfortunately, a book must be projected in a totally ordered way on the page axis, but that's not the way mathematics "is", so readers have to make choices how to reset certain topics in parallel for themselves, rather than in succession.

I have inserted cross-references to help them do this, but different people will make different choices at different times depending on different circumstances.

The book splits naturally into several parts. The first part introduces the basic notions of algebra. After these basic notions, the book splits in two major directions: the direction of algebraic equations including the Galois theory in Part II; and the direction of linear and multilinear algebra in Parts III and IV. There is some sporadic feedback between them, but their unification takes place at the next level of mathematics, which is suggested, for instance, in §15 of Chapter VI. Indeed, the study of algebraic extensions of the rationals can be carried out from two points of view which are complementary and interrelated: representing the Galois group of the algebraic closure in groups of matrices (the linear approach), and giving an explicit determination of the irrationalities generating algebraic extensions (the equations approach). At the moment, representations in GL_2 are at the center of attention from various quarters, and readers will see GL_2 appear several times throughout the book. For instance, I have found it appropriate to add a section describing all irreducible characters of $GL_2(F)$ when F is a finite field. Ultimately, GL_2 will appear as the simplest but typical case of groups of Lie types, occurring both in a differential context and over finite fields or more general arithmetic rings for arithmetic applications.

After almost a decade since the second edition, I find that the basic topics of algebra have become stable, with one exception. I have added two sections on elimination theory, complementing the existing section on the resultant. Algebraic geometry having progressed in many ways, it is now sometimes returning to older and harder problems, such as searching for the effective construction of polynomials vanishing on certain algebraic sets, and the older elimination procedures of last century serve as an introduction to those problems.

Except for this addition, the main topics of the book are unchanged from the second edition, but I have tried to improve the book in several ways.

First, some topics have been reordered. I was informed by readers and reviewers of the tension existing between having a textbook usable for relatively inexperienced students, and a reference book where results could easily be found in a systematic arrangement. I have tried to reduce this tension by moving all the homological algebra to a fourth part, and by integrating the commutative algebra with the chapter on algebraic sets and elimination theory, thus giving an introduction to different points of view leading toward algebraic geometry.

The book as a text and a reference

In teaching the course, one might wish to push into the study of algebraic equations through Part II, or one may choose to go first into the linear algebra of Parts III and IV. One semester could be devoted to each, for instance. The chapters have been so written as to allow maximal flexibility in this respect, and I have frequently committed the crime of lèse-Bourbaki by repeating short arguments or definitions to make certain sections or chapters logically independent of each other.

Granting the material which under no circumstances can be omitted from a basic course, there exist several options for leading the course in various directions. It is impossible to treat all of them with the same degree of thoroughness. The precise point at which one is willing to stop in any given direction will depend on time, place, and mood. However, any book with the aims of the present one must include a choice of topics, pushing ahead in deeper waters, while stopping short of full involvement.

There can be no universal agreement on these matters, not even between the author and himself. Thus the concrete decisions as to what to include and what not to include are finally taken on grounds of general coherence and aesthetic balance. Anyone teaching the course will want to impress their own personality on the material, and may push certain topics with more vigor than I have, at the expense of others. Nothing in the present book is meant to inhibit this.

Unfortunately, the goal to present a fairly comprehensive perspective on algebra required a substantial increase in size from the first to the second edition, and a moderate increase in this third edition. These increases require some decisions as to what to omit in a given course.

Many shortcuts can be taken in the presentation of the topics, which admits many variations. For instance, one can proceed into field theory and Galois theory immediately after giving the basic definitions for groups, rings, fields, polynomials in one variable, and vector spaces. Since the Galois theory gives very quickly an impression of depth, this is very satisfactory in many respects.

It is appropriate here to recall my original indebtedness to Artin, who first taught me algebra. The treatment of the basics of Galois theory is much influenced by the presentation in his own monograph.

Audience and background

As I already stated in the forewords of previous editions, the present book is meant for the graduate level, and I expect most of those coming to it to have had suitable exposure to some algebra in an undergraduate course, or to have appropriate mathematical maturity. I expect students taking a graduate course to have had some exposure to vector spaces, linear maps, matrices, and they will no doubt have seen polynomials at the very least in calculus courses.

My books *Undergraduate Algebra* and *Linear Algebra* provide more than enough background for a graduate course. Such elementary texts bring out in parallel the two basic aspects of algebra, and are organized differently from the present book, where both aspects are deepened. Of course, some aspects of the linear algebra in Part III of the present book are more "elementary" than some aspects of Part II, which deals with Galois theory and the theory of polynomial equations in several variables. Because Part II has gone deeper into the study of algebraic equations, of necessity the parallel linear algebra occurs only later in the total ordering of the book. Readers should view both parts as running simultaneously.

Unfortunately, the amount of algebra which one should ideally absorb during this first year in order to have a proper background (irrespective of the subject in which one eventually specializes) exceeds the amount which can be covered physically by a lecturer during a one-year course. Hence more material must be included than can actually be handled in class. I find it essential to bring this material to the attention of graduate students.

I hope that the various additions and changes make the book easier to use as a text. By these additions, I have tried to expand the general mathematical perspective of the reader, insofar as algebra relates to other parts of mathematics.

Acknowledgements

I am indebted to many people who have contributed comments and criticisms for the previous editions, but especially to Daniel Bump, Steven Krantz, and Diane Meuser, who provided extensive comments as editorial reviewers for Addison-Wesley. I found their comments very stimulating and valuable in preparing this third edition. I am much indebted to Barbara Holland for obtaining these reviews when she was editor. I am also indebted to Karl Matsumoto who supervised production under very trying circumstances. Finally I thank the many people who have made suggestions and corrections, especially George Bergman, Chee-Whye Chin, Ki-Bong Nam, David Wasserman, and Randy Scott, who provided me with a list of errata. I also thank Thomas Shiple and Paul Vojta for their lists of errata to the third edition. These have been corrected in the subsequent printings.

Serge Lang
New Haven

For the 2002 and beyond Springer printings

From now on, *Algebra* appears with Springer-Verlag, like the rest of my books. With this change, I considered the possibility of a new edition, but decided against it. I view the book as very stable. The only addition which I would make, if starting from scratch, would be some of the algebraic properties of SL_n and GL_n (over $\mathbf{R}$ or $\mathbf{C}$), beyond the proof of simplicity in Chapter XIII. As things stood, I just inserted some exercises concerning some aspects which everybody should know. Readers can see these worked out in Jorgenson/Lang, *Spherical Inversion on* $SL_n(\mathbf{R})$, Springer Verlag 2001, as well as other basic algebraic properties on which analysis is superimposed so that algebra in this context appears as a supporting tool.

I thank specifically Tom von Foerster, Ina Lindeman and Mark Spencer for their editorial support at Springer, as well as Terry Kornak and Brian Howe whc have taken care of production.

Serge Lang
New Haven 2002

Logical Prerequisites

We assume that the reader is familiar with sets, and with the symbols $\cap$, $\cup$, $\supset$, $\subset$, $\in$. If A, B are sets, we use the symbol $A \subset B$ to mean that A is contained in B but may be equal to B. Similarly for $A \supset B$.

If $f: A \to B$ is a mapping of one set into another, we write

$$x \mapsto f(x)$$

to denote the effect of f on an element x of A. We distinguish between the arrows $\to$ and $\mapsto$. We denote by $f(A)$ the set of all elements $f(x)$, with $x \in A$.

Let $f: A \to B$ be a mapping (also called a map). We say that f is **injective** if $x \neq y$ implies $f(x) \neq f(y)$. We say f is **surjective** if given $b \in B$ there exists $a \in A$ such that $f(a) = b$. We say that f is **bijective** if it is both surjective and injective.

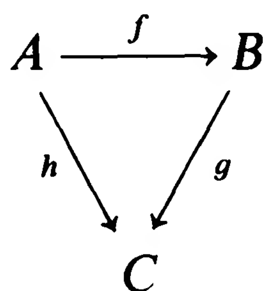
A subset A of a set B is said to be **proper** if $A \neq B$.

Let $f: A \to B$ be a map, and A' a subset of A. The restriction of f to A' is a map of A' into B denoted by $f|A'$.

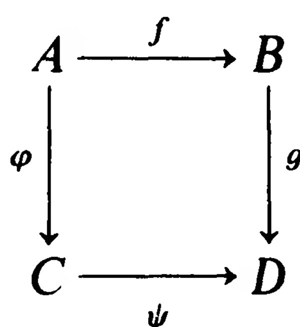
If $f: A \to B$ and $g: B \to C$ are maps, then we have a composite map $g \circ f$ such that $(g \circ f)(x) = g(f(x))$ for all $x \in A$.

Let $f: A \to B$ be a map, and B' a subset of B. By $f^{-1}(B')$ we mean the subset of A consisting of all $x \in A$ such that $f(x) \in B'$. We call it the **inverse image** of B'. We call $f(A)$ the **image** of f.

A **diagram**



is said to be **commutative** if $g \circ f = h$. Similarly, a **diagram**



is said to be **commutative** if $g \circ f = \psi \circ \varphi$. We deal sometimes with more complicated diagrams, consisting of arrows between various objects. Such diagrams are called commutative if, whenever it is possible to go from one object to another by means of two sequences of arrows, say

$$A_1 \xrightarrow{f_1} A_2 \xrightarrow{f_2} \cdots \xrightarrow{f_{n-1}} A_n$$

and

$$A_1 \xrightarrow{g_1} B_2 \xrightarrow{g_2} \cdots \xrightarrow{g_{m-1}} B_m = A_n,$$

then

$$f_{n-1} \circ \cdots \circ f_1 = g_{m-1} \circ \cdots \circ g_1,$$

in other words, the composite maps are equal. Most of our diagrams are composed of triangles or squares as above, and to verify that a diagram consisting of triangles or squares is commutative, it suffices to verify that each triangle and square in it is commutative.

We assume that the reader is acquainted with the integers and rational numbers, denoted respectively by **Z** and **Q**. For many of our examples, we also assume that the reader knows the real and complex numbers, denoted by **R** and **C**.

Let A and I be two sets. By a family of elements of A, indexed by I, one means a map $f: I \to A$. Thus for each $i \in I$ we are given an element $f(i) \in A$. Although a family does not differ from a map, we think of it as determining a collection of objects from A, and write it often as

$$\{f(i)\}_{i \in I}$$

or

$$\{a_i\}_{i \in I},$$

writing a_i instead of $f(i)$. We call I the indexing set.

We assume that the reader knows what an equivalence relation is. Let A be a set with an equivalence relation, let E be an equivalence class of elements of A. We sometimes try to define a map of the equivalence classes into some set B. To define such a map f on the class E, we sometimes first give its value on an element $x \in E$ (called a representative of E), and then show that it is independent of the choice of representative $x \in E$. In that case we say that f is **well defined**.

We have products of sets, say finite products $A \times B$, or $A_1 \times \cdots \times A_n$, and products of families of sets.

We shall use Zorn's lemma, which we describe in Appendix 2.

We let $\#(S)$ denote the number of elements of a set S, also called the **cardinality** of S. The notation is usually employed when S is finite. We also write $\#(S) = \text{card}(S)$.

CONTENTS

Part One The Basic Objects of Algebra

Part One

THE BASIC OBJECTS OF ALGEBRA

This part introduces the basic notions of algebra, and the main difficulty for the beginner is to absorb a reasonable vocabulary in a short time. None of the concepts is difficult, but there is an accumulation of new concepts which may sometimes seem heavy.

To understand the next parts of the book, the reader needs to know essentially only the basic definitions of this first part. Of course, a theorem may be used later for some specific and isolated applications, but on the whole, we have avoided making long logical chains of interdependence.

CHAPTER I

Groups

§1. MONOIDS

Let S be a set. A mapping

$$S \times S \to S$$

is sometimes called a **law of composition** (of S into itself). If x, y are elements of S, the image of the pair (x, y) under this mapping is also called their **product** under the law of composition, and will be denoted by xy. (Sometimes, we also write $x \cdot y$, and in many cases it is also convenient to use an additive notation, and thus to write $x + y$. In that case, we call this element the **sum** of x and y. It is customary to use the notation $x + y$ only when the relation $x + y = y + x$ holds.)

Let S be a set with a law of composition. If x, y, z are elements of S, then we may form their product in two ways: $(xy)z$ and $x(yz)$. If $(xy)z = x(yz)$ for all x, y, z in S then we say that the law of composition is **associative**.

An element e of S such that $ex = x = xe$ for all $x \in S$ is called a **unit element**. (When the law of composition is written additively, the unit element is denoted by 0, and is called a **zero element**.) A unit element is unique, for if e' is another unit element, we have

$$e = ee' = e'$$

by assumption. In most cases, the unit element is written simply 1 (instead of e). For most of this chapter, however, we shall write e so as to avoid confusion in proving the most basic properties.

A **monoid** is a set G, with a law of composition which is associative, and having a unit element (so that in particular, G is not empty).

Let G be a monoid, and $x_1, \ldots, x_n$ elements of G (where n is an integer > 1). We define their product inductively:

$$\prod_{\nu=1}^{n} x_\nu = x_1 \cdots x_n = (x_1 \cdots x_{n-1})x_n.$$

We then have the following rule:

$$\prod_{\mu=1}^{m} x_\mu \cdot \prod_{\nu=1}^{n} x_{m+\nu} = \prod_{\nu=1}^{m+n} x_\nu,$$

which essentially asserts that we can insert parentheses in any manner in our product without changing its value. The proof is easy by induction, and we shall leave it as an exercise.

One also writes

$$\prod_{m+1}^{m+n} x_\nu \quad \text{instead of} \quad \prod_{\nu=1}^{n} x_{m+\nu}$$

and we define

$$\prod_{\nu=1}^{0} x_\nu = e.$$

As a matter of convention, we agree also that the empty product is equal to the unit element.

It would be possible to define more general laws of composition, i.e. maps $S_1 \times S_2 \to S_3$ using arbitrary sets. One can then express associativity and commutativity in any setting for which they make sense. For instance, for commutativity we need a law of composition

$$f: S \times S \to T$$

where the two sets of departure are the same. **Commutativity** then means $f(x, y) = f(y, x)$, or $xy = yx$ if we omit the mapping f from the notation. For associativity, we leave it to the reader to formulate the most general combination of sets under which it will work. We shall meet special cases later, for instance arising from maps

$$S \times S \to S \quad \text{and} \quad S \times T \to T.$$

Then a product $(xy)z$ makes sense with $x \in S$, $y \in S$, and $z \in T$. The product $x(yz)$ also makes sense for such elements x, y, z and thus it makes sense to say that our law of composition is associative, namely to say that for all x, y, z as above we have $(xy)z = x(yz)$.

If the law of composition of G is commutative, we also say that G is **commutative (or abelian)**.

Let G be a commutative monoid, and $x_1, \ldots, x_n$ elements of G. Let ψ be a bijection of the set of integers $(1, \ldots, n)$ onto itself. Then

$$\prod_{\nu=1}^{n} x_{\psi(\nu)} = \prod_{\nu=1}^{n} x_\nu.$$

We prove this by induction, it being obvious for $n = 1$. We assume it for $n - 1$. Let k be an integer such that $\psi(k) = n$. Then

$$\prod_{1}^{n} x_{\psi(\nu)} = \prod_{1}^{k-1} x_{\psi(\nu)} \cdot x_{\psi(k)} \cdot \prod_{1}^{n-k} x_{\psi(k+\nu)}$$

$$= \prod_{1}^{k-1} x_{\psi(\nu)} \cdot \prod_{1}^{n-k} x_{\psi(k+\nu)} \cdot x_{\psi(k)}.$$

Define a map φ of $(1, \ldots, n - 1)$ into itself by the rule

$$\varphi(\nu) = \psi(\nu) \qquad \text{if} \quad \nu < k,$$

$$\varphi(\nu) = \psi(\nu + 1) \qquad \text{if} \quad \nu \geqq k.$$

Then

$$\prod_{1}^{n} x_{\psi(\nu)} = \prod_{1}^{k-1} x_{\varphi(\nu)} \prod_{1}^{n-k} x_{\varphi(k-1+\nu)} \cdot x_n$$

$$= \prod_{1}^{n-1} x_{\varphi(\nu)} \cdot x_n,$$

which, by induction, is equal to $x_1 \cdots x_n$, as desired.

Let G be a commutative monoid, let I be a set, and let $f : I \to G$ be a mapping such that $f(i) = e$ for almost all $i \in I$. (Here and thereafter, **almost all** will mean *all but a finite number*.) Let I_0 be the subset of I consisting of those i such that $f(i) \neq e$. By

$$\prod_{i \in I} f(i)$$

we shall mean the product

$$\prod_{i \in I_0} f(i)$$

taken in any order (the value does not depend on the order, according to the preceding remark). It is understood that the empty product is equal to e.

When G is written additively, then instead of a product sign, we write the sum sign Σ.

There are a number of formal rules for dealing with products which it would be tedious to list completely. We give one example. Let I, J be two sets, and

$f: I \times J \to G$ a mapping into a commutative monoid which takes the value e for almost all pairs (i, j). Then

$$\prod_{i \in I} \left[\prod_{j \in J} f(i, j) \right] = \prod_{j \in J} \left[\prod_{i \in I} f(i, j) \right].$$

We leave the proof as an exercise.

As a matter of notation, we sometimes write $\prod f(i)$, omitting the signs $i \in I$, if the reference to the indexing set is clear.

Let x be an element of a monoid G. For every integer $n \geqq 0$ we define x^n to be

$$\prod_{1}^{n} x,$$

so that in particular we have $x^0 = e$, $x^1 = x$, $x^2 = xx, \ldots$. We obviously have $x^{(n+m)} = x^n x^m$ and $(x^n)^m = x^{nm}$. Furthermore, from our preceding rules of associativity and commutativity, if x, y are elements of G such that $xy = yx$, then $(xy)^n = x^n y^n$. We leave the formal proof as an exercise.

If S, S' are two subsets of a monoid G, then we define SS' to be the subset consisting of all elements xy, with $x \in S$ and $y \in S'$. Inductively, we can define the product of a finite number of subsets, and we have associativity. For instance, if S, S', S'' are subsets of G, then $(SS')S'' = S(S'S'')$. Observe that $GG = G$ (because G has a unit element). If $x \in G$, then we define xS to be $\{x\}S$, where $\{x\}$ is the set consisting of the single element x. Thus xS consists of all elements xy, with $y \in S$.

By a **submonoid** of G, we shall mean a subset H of G containing the unit element e, and such that, if $x, y \in H$ then $xy \in H$ (we say that H is **closed** under the law of composition). It is then clear that H is itself a monoid, under the law of composition induced by that of G.

If x is an element of a monoid G, then the subset of powers x^n $(n = 0, 1, \ldots)$ is a submonoid of G.

The set of integers $\geqq 0$ under addition is a monoid.

Later we shall define rings. If R is a commutative ring, we shall deal with multiplicative subsets S, that is subsets containing the unit element, and such that if $x, y \in S$ then $xy \in S$. Such subsets are monoids.

A routine example. Let $\mathbf{N}$ be the natural numbers, i.e. the integers $\geqq 0$. Then $\mathbf{N}$ is an additive monoid. In some applications, it is useful to deal with a multiplicative version. See the definition of polynomials in Chapter II, §3, where a higher-dimensional version is also used for polynomials in several variables.

An interesting example. We assume that the reader is familiar with the terminology of elementary topology. Let M be the set of homeomorphism classes of compact (connected) surfaces. We shall define an addition in M. Let S, S' be compact surfaces. Let D be a small disc in S, and D' a small disc in S'. Let C, C' be the circles which form the boundaries of D and D' respectively. Let D_0, D'_0 be the interiors of D and D' respectively, and glue $S - D_0$ to $S' - D'_0$ by identifying C with C'. It can be shown that the resulting surface is independent,

up to homeomorphism, of the various choices made in the preceding construction. If σ, σ' denote the homeomorphism classes of S and S' respectively, we define $\sigma + \sigma'$ to be the class of the surface obtained by the preceding gluing process. It can be shown that this addition defines a monoid structure on M, whose unit element is the class of the ordinary 2-sphere. Furthermore, if τ denotes the class of the torus, and π denotes the class of the projective plane, then every element σ of M has a unique expression of the form

$$\sigma = n\tau + m\pi$$

where n is an integer $\geqq 0$ and $m = 0, 1$, or 2. We have $3\pi = \tau + \pi$.

(The reasons for inserting the preceding example are twofold: First to relieve the essential dullness of the section. Second to show the reader that monoids exist in nature. Needless to say, the example will not be used in any way throughout the rest of the book.)

Still other examples. At the end of Chapter III, §4, we shall remark that isomorphism classes of modules over a ring form a monoid under the direct sum. In Chapter XV, §1, we shall consider a monoid consisting of equivalence classes of quadratic forms.

§2. GROUPS

A **group** G is a monoid, such that for every element $x \in G$ there exists an element $y \in G$ such that $xy = yx = e$. Such an element y is called an **inverse** for x. Such an inverse is unique, because if y' is also an inverse for x, then

$$y' = y'e = y'(xy) = (y'x)y = ey = y.$$

We denote this inverse by x^{-1} (or by $-x$ when the law of composition is written additively).

For any positive integer n, we let $x^{-n} = (x^{-1})^n$. Then the usual rules for exponentiation hold for all integers, not only for integers $\geqq 0$ (as we pointed out for monoids in §1). The trivial proofs are left to the reader.

In the definitions of unit elements and inverses, we could also define left units and left inverses (in the obvious way). One can easily prove that these are also units and inverses respectively under suitable conditions. Namely:

Let G be a set with an associative law of composition, let e be a left unit for that law, and assume that every element has a left inverse. Then e is a unit, and each left inverse is also an inverse. In particular, G is a group.

To prove this, let $a \in G$ and let $b \in G$ be such that $ba = e$. Then

$$bab = eb = b.$$

Multiplying on the left by a left inverse for b yields

$$ab = e,$$

or in other words, b is also a right inverse for a. One sees also that a is a left

inverse for b. Furthermore,

$$ae = aba = ea = a,$$

whence e is a right unit.

Example. Let G be a group and S a nonempty set. The set of maps $M(S, G)$ is itself a group; namely for two maps f, g of S into G we define fg to be the map such that

$$(fg)(x) = f(x)g(x),$$

and we define f^{-1} to be the map such that $f^{-1}(x) = f(x)^{-1}$. It is then trivial to verify that $M(S, G)$ is a group. If G is commutative, so is $M(S, G)$, and when the law of composition in G is written additively, so is the law of composition in $M(S, G)$, so that we would write $f + g$ instead of fg, and $-f$ instead of f^{-1}.

Example. Let S be a non-empty set. Let G be the set of bijective mappings of S onto itself. Then G is a group, the law of composition being ordinary composition of mappings. The unit element of G is the identity map of S, and the other group properties are trivially verified. The elements of G are called **permutations** of S. We also denote G by Perm(S). For more information on Perm(S) when S is finite, see §5 below.

Example. Let us assume here the basic notions of linear algebra. Let k be a field and V a vector space over k. Let $GL(V)$ denote the set of invertible k-linear maps of V onto itself. Then $GL(V)$ is a group under composition of mappings. Similarly, let k be a field and let $GL(n, k)$ be the set of invertible $n \times n$ matrices with components in k. Then $GL(n, k)$ is a group under the multiplication of matrices. For $n \geqq 2$, this group is not commutative.

Example. The group of automorphisms. We recommend that the reader now refer immediately to §11, where the notion of a category is defined, and where several examples are given. For any object A in a category, its automorphisms form a group denoted by Aut(A). Permutations of a set and the linear automorphisms of a vector space are merely examples of this more general structure.

Example. The set of rational numbers forms a group under addition. The set of non-zero rational numbers forms a group under multiplication. Similar statements hold for the real and complex numbers.

Example. Cyclic groups. The integers **Z** form an additive group. A group is defined to be **cyclic** if there exists an element $a \in G$ such that every element of G (written multiplicatively) is of the form a^n for some integer n. If G is written additively, then every element of a cyclic group is of the form na. One calls a a **cyclic generator**. Thus **Z** is an additive cyclic group with generator 1, and also with generator -1. There are no other generators. Given a positive integer n, the n-th roots of unity in the complex numbers form a cyclic group of order n. In terms of the usual notation, $e^{2\pi i/n}$ is a generator for this group. So is $e^{2\pi ir/n}$

with $r \in \mathbf{Z}$ and r prime to n. A generator for this group is called a **primitive** n-th root of unity.

Example. The direct product. Let G_1, G_2 be groups. Let $G_1 \times G_2$ be the direct product as sets, so $G_1 \times G_2$ is the set of all pairs (x_1, x_2) with $x_i \in G_i$. We define the product componentwise by

$$(x_1, x_2)(y_1, y_2) = (x_1 y_1, x_2 y_2).$$

Then $G_1 \times G_2$ is a group, whose unit element is (e_1, e_2) (where e_i is the unit element of G_i). Similarly, for n groups we define $G_1 \times \cdots \times G_n$ to be the set of n-tuples with $x_i \in G_i$ $(i = 1, \ldots, n)$, and componentwise multiplication. Even more generally, let I be a set, and for each $i \in I$, let G_i be a group. Let $G = \prod G_i$ be the set-theoretic product of the sets G_i. Then G is the set of all families $(x_i)_{i \in I}$ with $x_i \in G_i$. We can define a group structure on G by componentwise multiplication, namely, if $(x_i)_{i \in I}$ and $(y_i)_{i \in I}$ are two elements of G, we define their product to be $(x_i y_i)_{i \in I}$. We define the inverse of $(x_i)_{i \in I}$ to be $(x_i^{-1})_{i \in I}$. It is then obvious that G is a group called the **direct product** of the family.

Let G be a group. A **subgroup** H of G is a subset of G containing the unit element, and such that H is closed under the law of composition and inverse (i.e. it is a submonoid, such that if $x \in H$ then $x^{-1} \in H$). A subgroup is called **trivial** if it consists of the unit element alone. The intersection of an arbitrary non-empty family of subgroups is a subgroup (trivial verification).

Let G be a group and S a subset of G. We shall say that S **generates** G, or that S is a set of **generators** for G, if every element of G can be expressed as a product of elements of S or inverses of elements of S, i.e. as a product $x_1 \cdots x_n$ where each x_i or x_i^{-1} is in S. It is clear that the set of all such products is a subgroup of G (the empty product is the unit element), and is the smallest subgroup of G containing S. Thus S generates G if and only if the smallest subgroup of G containing S is G itself. If G is generated by S, then we write $G = \langle S \rangle$. By definition, a cyclic group is a group which has one generator. Given elements $x_1, \ldots, x_n \in G$, these elements generate a subgroup $\langle x_1, \ldots, x_n \rangle$, namely the set of all elements of G of the form

$$x_{i_1}^{k_1} \cdots x_{i_r}^{k_r} \quad \text{with} \quad k_1, \ldots, k_r \in \mathbf{Z}.$$

A single element $x \in G$ generates a cyclic subgroup.

Example. There are two non-abelian groups of order 8. One is the **group of symmetries of the square**, generated by two elements σ, τ such that

$$\sigma^4 = \tau^2 = e \quad \text{and} \quad \tau\sigma\tau^{-1} = \sigma^3.$$

The other is the **quaternion group**, generated by two elements, i, j such that if we put $k = ij$ and $m = i^2$, then

$$i^4 = j^4 = k^4 = e, \quad i^2 = j^2 = k^2 = m, \quad ij = mji.$$

After you know enough facts about groups, you can easily do Exercise 35.

Let G, G' be monoids. A **monoid-homomorphism** (or simply **homomorphism**) of G into G' is a mapping $f: G \to G'$ such that $f(xy) = f(x)f(y)$ for all $x, y \in G$, and mapping the unit element of G into that of G'. If G, G' are groups, a **group-homomorphism** of G into G' is simply a monoid-homomorphism.

We sometimes say: "Let $f: G \to G'$ be a group-homomorphism" to mean: "Let G, G' be groups, and let f be a homomorphism from G into G'."

Let $f: G \to G'$ be a group-homomorphism. Then

$$f(x^{-1}) = f(x)^{-1}$$

because if e, e' are the unit elements of G, G' respectively, then

$$e' = f(e) = f(xx^{-1}) = f(x)f(x^{-1}).$$

Furthermore, if G, G' are groups and $f: G \to G'$ is a map such that

$$f(xy) = f(x)f(y)$$

for all x, y in G, then $f(e) = e'$ because $f(ee) = f(e)$ and also $= f(e)f(e)$. Multiplying by the inverse of $f(e)$ shows that $f(e) = e'$.

Let G, G' be monoids. A homomorphism $f: G \to G'$ is called an **isomorphism** if there exists a homomorphism $g: G' \to G$ such that $f \circ g$ and $g \circ f$ are the identity mappings (in G' and G respectively). It is trivially verified that f is an isomorphism if and only if f is bijective. The existence of an isomorphism between two groups G and G' is sometimes denoted by $G \approx G'$. If $G = G'$, we say that isomorphism is an **automorphism**. A homomorphism of G into itself is also called an **endomorphism**.

Example. Let G be a monoid and x an element of G. Let $\mathbf{N}$ denote the (additive) monoid of integers $\geqq 0$. Then the map $f: \mathbf{N} \to G$ such that $f(n) = x^n$ is a homomorphism. If G is a group, we can extend f to a homomorphism of $\mathbf{Z}$ into G (x^n is defined for all $n \in \mathbf{Z}$, as pointed out previously). The trivial proofs are left to the reader.

Let n be a fixed integer and let G be a *commutative* group. Then one verifies easily that the map

$$x \mapsto x^n$$

from G into itself is a homomorphism. So is the map $x \mapsto x^{-1}$. The map $x \mapsto x^n$ is called the n-th **power map**.

Example. Let $I = \{i\}$ be an indexing set, and let $\{G_i\}$ be a family of groups. Let $G = \prod G_i$ be their direct product. Let

$$p_i \colon G \to G_i$$

be the projection on the i-th factor. Then p_i is a homomorphism.

Let G be a group, S a set of generators for G, and G' another group. Let $f: S \to G'$ be a map. If there exists a homomorphism $\bar{f}$ of G into G' whose restriction to S is f, then there is only one.

In other words, f has at most one extension to a homomorphism of G into G'. This is obvious, but will be used many times in the sequel.

Let $f: G \to G'$ and $g: G' \to G''$ be two group-homomorphisms. Then the composite map $g \circ f$ is a group-homomorphism. If f, g are isomorphisms then so is $g \circ f$. Furthermore $f^{-1}: G' \to G$ is also an isomorphism. In particular, the set of all automorphisms of G is itself a group, denoted by Aut(G).

Let $f: G \to G'$ be a group-homomorphism. Let e, e' be the respective unit elements of G, G'. We define the **kernel** of f to be the subset of G consisting of all x such that $f(x) = e'$. From the definitions, it follows at once that the kernel H of f is a subgroup of G. (Let us prove for instance that H is closed under the inverse mapping. Let $x \in H$. Then

$$f(x^{-1})f(x) = f(e) = e'.$$

Since $f(x) = e'$, we have $f(x^{-1}) = e'$, whence $x^{-1} \in H$. We leave the other verifications to the reader.)

Let $f: G \to G'$ be a group-homomorphism again. Let H' be the **image** of f. Then H' is a subgroup of G', because it contains e', and if $f(x), f(y) \in H'$, then $f(xy) = f(x)f(y)$ lies also in H'. Furthermore, $f(x^{-1}) = f(x)^{-1}$ is in H', and hence H' is a subgroup of G'.

The kernel and image of f are sometimes denoted by Ker f and Im f.

A homomorphism $f: G \to G'$ which establishes an isomorphism between G and its image in G' will also be called an **embedding**.

A homomorphism whose kernel is trivial is injective.

To prove this, suppose that the kernel of f is trivial, and let $f(x) = f(y)$ for some $x, y \in G$. Multiplying by $f(y^{-1})$ we obtain

$$f(xy^{-1}) = f(x)f(y^{-1}) = e'.$$

Hence xy^{-1} lies in the kernel, hence $xy^{-1} = e$, and $x = y$. If in particular f is also surjective, then f is an isomorphism. Thus a surjective homomorphism whose kernel is trivial must be an isomorphism. We note that an injective homomorphism is an embedding.

An injective homomorphism is often denoted by a special arrow, such as

$$f: G \hookrightarrow G'.$$

There is a useful criterion for a group to be a direct product of subgroups:

Proposition 2.1. *Let G be a group and let H, K be two subgroups such that $H \cap K = e$, $HK = G$, and such that $xy = yx$ for all $x \in H$ and $y \in K$. Then the map*

$$H \times K \to G$$

such that $(x, y) \mapsto xy$ is an isomorphism.

Proof. It is obviously a homomorphism, which is surjective since $HK = G$.

If (x, y) is in its kernel, then $x = y^{-1}$, whence x lies in both H and K, and $x = e$, so that $y = e$ also, and our map is an isomorphism.

We observe that Proposition 2.1 generalizes by induction to a finite number of subgroups $H_1, \ldots, H_n$ whose elements commute with each other, such that

$$H_1 \cdots H_n = G,$$

and such that

$$H_{i+1} \cap (H_1 \cdots H_i) = e.$$

In that case, G is isomorphic to the direct product

$$H_1 \times \cdots \times H_n.$$

Let G be a group and H a subgroup. A **left coset** of H in G is a subset of G of type aH, for some element a of G. An element of aH is called a **coset representative** of aH. The map $x \mapsto ax$ induces a bijection of H onto aH. Hence any two left cosets have the same cardinality.

Observe that if a, b are elements of G and aH, bH are cosets having one element in common, then they are equal. Indeed, let $ax = by$ with $x, y \in H$. Then $a = byx^{-1}$. But $yx^{-1} \in H$. Hence $aH = b(yx^{-1})H = bH$, because for any $z \in H$ we have $zH = H$.

We conclude that G is the disjoint union of the left cosets of H. A similar remark applies to **right cosets** (i.e. subsets of G of type Ha). The number of left cosets of H in G is denoted by $(G : H)$, and is called the (left) **index** of H in G. The index of the trivial subgroup is called the **order** of G and is written $(G : 1)$. From the above conclusion, we get:

Proposition 2.2. *Let G be a group and H a subgroup. Then*

$$(G : H)(H : 1) = (G : 1),$$

in the sense that if two of these indices are finite, so is the third and equality holds as stated. If $(G : 1)$ is finite, the order of H divides the order of G.

More generally, let H, K be subgroups of G and let $H \supset K$. Let $\{x_i\}$ be a set of (left) coset representatives of K in H and let $\{y_j\}$ be a set of coset representatives of H in G. Then we contend that $\{y_j x_i\}$ is a set of coset representatives of K in G.

Proof. Note that

$$H = \bigcup_i x_i K \qquad \text{(disjoint)},$$

$$G = \bigcup_j y_j H \qquad \text{(disjoint)}.$$

Hence

$$G = \bigcup_{i,j} y_j x_i K.$$

We must show that this union is disjoint, i.e. that the $y_j x_i$ represent distinct cosets. Suppose

$$y_j x_i K = y_{j'} x_{i'} K$$

for a pair of indices (j, i) and (j', i'). Multiplying by H on the right, and noting that x_i, $x_{i'}$ are in H, we get

$$y_j H = y_{j'} H,$$

whence $y_j = y_{j'}$. From this it follows that $x_i K = x_{i'} K$ and therefore that $x_i = x_{i'}$, as was to be shown.

The formula of Proposition 2.2 may therefore be generalized by writing

$$(G : K) = (G : H)(H : K),$$

with the understanding that if two of the three indices appearing in this formula are finite, then so is the third and the formula holds.

The above results are concerned systematically with left cosets. For the right cosets, see Exercise 10.

Example. A group of prime order is cyclic. Indeed, let G have order p and let $a \in G$, $a \neq e$. Let H be the subgroup generated by a. Then $\#(H)$ divides p and is $\neq 1$, so $\#(H) = p$ and so $H = G$, which is therefore cyclic.

Example. Let $J_n = \{1, \ldots, n\}$. Let S_n be the group of permutations of J_n. We define a **transposition** to be a permutation τ such that there exist two elements $r \neq s$ in J_n for which $\tau(r) = s$, $\tau(s) = r$, and $\tau(k) = k$ for all $k \neq r, s$. Note that the transpositions generate S_n. Indeed, say σ is a permutation, $\sigma(n) = k \neq n$. Let τ be the transposition interchanging k, n. Then $\tau\sigma$ leaves n fixed, and by induction, we can write $\tau\sigma$ as a product of transpositions in $\text{Perm}(J_{n-1})$, thus proving that transpositions generate S_n.

Next we note that $\#(S_n) = n!$. Indeed, let H be the subgroup of S_n consisting of those elements which leave n fixed. Then H may be identified with S_{n-1}. If σ_i $(i = 1, \ldots, n)$ is an element of S_n such that $\sigma_i(n) = i$, then it is immediately verified that $\sigma_1, \ldots, \sigma_n$ are coset representatives of H. Hence by induction

$$(S_n : 1) = n(H : 1) = n!.$$

Observe that for σ_i we could have taken the transposition τ_i, which interchanges i and n (except for $i = n$, where we could take σ_n to be the identity).

§3. NORMAL SUBGROUPS

We have already observed that the kernel of a group-homomorphism is a subgroup. We now wish to characterize such subgroups.

Let $f : G \to G'$ be a group-homomorphism, and let H be its kernel. If x is an element of G, then $xH = Hx$, because both are equal to $f^{-1}(f(x))$. We can also rewrite this relation as $xHx^{-1} = H$.

Conversely, let G be a group, and let H be a subgroup. Assume that for all elements x of G we have $xH \subset Hx$ (or equivalently, $xHx^{-1} \subset H$). If we write x^{-1} instead of x, we get $H \subset xHx^{-1}$, whence $xHx^{-1} = H$. Thus our condition is equivalent to the condition $xHx^{-1} = H$ for all $x \in G$. A subgroup H satisfying this condition will be called **normal**. We shall now see that a normal subgroup is the kernel of a homomorphism.

Let G' be the set of cosets of H. (By assumption, a left coset is equal to a right coset, so we need not distinguish between them.) If xH and yH are cosets, then their product $(xH)(yH)$ is also a coset, because

$$xHyH = xyHH = xyH.$$

By means of this product, we have therefore defined a law of composition on G' which is associative. It is clear that the coset H itself is a unit element for this law of composition, and that $x^{-1}H$ is an inverse for the coset xH. Hence G' is a group.

Let $f: G \to G'$ be the mapping such that $f(x)$ is the coset xH. Then f is clearly a homomorphism, and (the subgroup) H is contained in its kernel. If $f(x) = H$, then $xH = H$. Since H contains the unit element, it follows that $x \in H$. Thus H is equal to the kernel, and we have obtained our desired homomorphism.

The group of cosets of a normal subgroup H is denoted by G/H (which we read G modulo H, or G mod H). The map f of G onto G/H constructed above is called the **canonical map**, and G/H is called the **factor group** of G by H.

Remarks

1. Let $\{H_i\}_{i \in I}$ be a family of normal subgroups of G. Then the subgroup

$$H = \bigcap_{i \in I} H_i$$

is a normal subgroup. Indeed, if $y \in H$, and $x \in G$, then xyx^{-1} lies in each H_i, whence in H.

2. Let S be a subset of G and let $N = N_S$ be the set of all elements $x \in G$ such that $xSx^{-1} = S$. Then N is obviously a subgroup of G, called the **normalizer** of S. If S consists of one element a, then N is also called the **centralizer** of a. More generally, let Z_S be the set of all elements $x \in G$ such that $xyx^{-1} = y$ for all $y \in S$. Then Z_S is called the **centralizer** of S. The centralizer of G itself is called the **center** of G. It is the subgroup of G consisting of all elements of G commuting with all other elements, and is obviously a normal subgroup of G.

Examples. We shall give more examples of normal subgroups later when we have more theorems to prove the normality. Here we give only two examples.

First, from linear algebra, note that the determinant is a homomorphism from the multiplicative group of square matrices into the multiplicative group of a field. The kernel is called the **special linear group**, and is normal.

Second, let G be the set of all maps $T_{a,b}$: $\mathbf{R} \to \mathbf{R}$ such that $T_{a,b}(x) = ax + b$, with $a \neq 0$ and b arbitrary. Then G is a group under composition of mappings. Let A be the multiplicative group of maps of the form $T_{a,0}$ (isomorphic to $\mathbf{R}^*$, the non-zero elements of $\mathbf{R}$), and let N be the group of translations $T_{1,b}$ with $b \in \mathbf{R}$. Then the reader will verify at once that $T_{a,b} \mapsto a$ is a homomorphism of G onto the multiplicative group, whose kernel is the group of translations, which is therefore normal. Furthermore, we have $G = AN = NA$, and $N \cap A = \{\text{id}\}$. In the terminology of Exercise 12, G is the **semidirect product** of A and N.

Let H be a subgroup of G. Then H is obviously a normal subgroup of its normalizer N_H. We leave the following statements as exercises:

If K is any subgroup of G containing H and such that H is normal in K, then $K \subset N_H$.

If K is a subgroup of N_H, then KH is a group and H is normal in KH.
The normalizer of H is the largest subgroup of G in which H is normal.

Let G be a group and H a normal subgroup. Let $x, y \in G$. We shall write

$$x \equiv y \pmod{H}$$

if x and y lie in the same coset of H, or equivalently if xy^{-1} (or $y^{-1}x$) lie in H. We read this relation "x and y are congruent modulo H."

When G is an additive group, then

$$x \equiv 0 \pmod{H}$$

means that x lies in H, and

$$x \equiv y \pmod{H}$$

means that $x - y$ (or $y - x$) lies in H. This notation of congruence is used mostly for additive groups.

Let

$$G' \xrightarrow{f} G \xrightarrow{g} G''$$

be a sequence of homomorphisms. We shall say that this sequence is **exact** if $\operatorname{Im} f = \operatorname{Ker} g$. For example, if H is a normal subgroup of G then the sequence

$$H \xrightarrow{j} G \xrightarrow{\varphi} G/H$$

is exact (where j = inclusion and φ = canonical map). A sequence of homomorphisms having more than one term, like

$$G_1 \xrightarrow{f_1} G_2 \xrightarrow{f_2} G_3 \to \cdots \xrightarrow{f_{n-1}} G_n,$$

is called **exact** if it is exact at each joint, i.e. if

$$\operatorname{Im} f_i = \operatorname{Ker} f_{i+1}$$

for each $i = 1, \ldots, n - 2$. For example to say that

$$0 \to G' \xrightarrow{f} G \xrightarrow{g} G'' \to 0$$

is exact means that f is injective, that $\operatorname{Im} f = \operatorname{Ker} g$, and that g is surjective. If $H = \operatorname{Ker} g$ then this sequence is essentially the same as the exact sequence

$$0 \to H \to G \to G/H \to 0.$$

More precisely, there exists a commutative diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & G' & \xrightarrow{f} & G & \xrightarrow{g} & G'' & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & H & \longrightarrow & G & \longrightarrow & G/H & \longrightarrow & 0 \end{array}$$

in which the vertical maps are isomorphisms, and the rows are exact.

Next we describe some homomorphisms, all of which are called **canonical**.

(i) Let G, G' be groups and $f: G \to G'$ a homomorphism whose kernel is H. Let $\varphi: G \to G/H$ be the canonical map. Then there exists a unique homomorphism $f_*: G/H \to G'$ such that $f = f_* \circ \varphi$, and f_* is injective.

To define f_*, let xH be a coset of H. Since $f(xy) = f(x)$ for all $y \in H$, we define $f_*(xH)$ to be $f(x)$. This value is independent of the choice of coset representative x, and it is then trivially verified that f_* is a homomorphism, is injective, and is the unique homomorphism satisfying our requirements. We shall say that f_* is **induced** by f.

Our homomorphism f_* induces an isomorphism

$$\lambda: G/H \to \operatorname{Im} f$$

of G/H onto the image of f, and thus f can be factored into the following succession of homomorphisms:

$$G \xrightarrow{\varphi} G/H \xrightarrow{\lambda} \operatorname{Im} f \xrightarrow{j} G'.$$

Here, j is the inclusion of $\operatorname{Im} f$ in G'.

(ii) Let G be a group and H a subgroup. Let N be the intersection of all normal subgroups containing H. Then N is normal, and hence is the smallest normal subgroup of G containing H. Let $f: G \to G'$ be a homomorphism whose kernel contains H. Then the kernel of f contains N, and there exists a unique homomorphism $f_*: G/N \to G'$, said to be induced by f, making the following diagram commutative:

$$\begin{array}{ccc} G & \xrightarrow{f} & G' \\ & {\scriptstyle\varphi}\searrow \quad \nearrow{\scriptstyle f_*} & \\ & G/N & \end{array}$$

As before, φ is the canonical map.

We can define f_* as in (1) by the rule

$$f_*(xN) = f(x).$$

This is well defined, and is trivially verified to satisfy all our requirements.

(iii) Let G be group and $H \supset K$ two normal subgroups of G. Then K is normal in H, and we can define a map of G/K onto G/H by associating with each coset xK the coset xH. It is immediately verified that this map is a homomorphism, and that its kernel consists of all cosets xK such that $x \in H$. Thus we have a canonical isomorphism

$$\boxed{(G/K)/(H/K) \approx G/H.}$$

One could also describe this isomorphism using (i) and (ii). We leave it to the reader to show that we have a commutative diagram

$$\begin{array}{ccccccccc}
0 & \longrightarrow & H & \longrightarrow & G & \longrightarrow & G/H & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle \text{can}} & & \downarrow{\scriptstyle \text{can}} & & \downarrow{\scriptstyle \text{id}} & & \\
0 & \longrightarrow & H/K & \longrightarrow & G/K & \longrightarrow & G/H & \longrightarrow & 0
\end{array}$$

where the rows are exact.

(iv) Let G be a group and let H, K be two subgroups. Assume that H is contained in the normalizer of K. Then $H \cap K$ is obviously a normal subgroup of H, and equally obviously $HK = KH$ is a subgroup of G. There is a surjective homomorphism

$$H \to HK/K$$

associating with each $x \in H$ the coset xK of K in the group HK. The reader will verify at once that the kernel of this homomorphism is exactly $H \cap K$. Thus we have a canonical isomorphism

$$\boxed{H/(H \cap K) \approx HK/K.}$$

(v) Let $f: G \to G'$ be a group homomorphism, let H' be a normal subgroup of G', and let $H = f^{-1}(H')$.

$$\begin{array}{ccc}
G & \longrightarrow & G' \\
\uparrow & & \uparrow \\
f^{-1}(H') & \longrightarrow & H'
\end{array}$$

Then $f^{-1}(H')$ is normal in G. [*Proof*: If $x \in G$, then $f(xHx^{-1}) = f(x)f(H)f(x)^{-1}$ is contained in H', so $xHx^{-1} \subset H$.] We then obtain a homomorphism

$$G \to G' \to G'/H'$$

composing f with the canonical map of G' onto G'/H', and the kernel of this composite is H. Hence we get an injective homomorphism

$$\bar{f}: G/H \to G'/H'$$

again called canonical, giving rise to the commutative diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & H & \longrightarrow & G & \longrightarrow & G/H & \longrightarrow & 0 \\ & & \downarrow & & \downarrow f & & \downarrow \bar{f} & & \\ 0 & \longrightarrow & H' & \longrightarrow & G' & \longrightarrow & G'/H' & \longrightarrow & 0. \end{array}$$

If f is surjective, then $\bar{f}$ is an isomorphism.

We shall now describe some applications of our homomorphism statements.

Let G be a group. A sequence of subgroups

$$G = G_0 \supset G_1 \supset G_2 \supset \cdots \supset G_m$$

is called a **tower** of subgroups. The tower is said to be **normal** if each G_{i+1} is normal in G_i ($i = 0, \dots, m-1$). It is said to be **abelian** (resp. **cyclic**) if it is normal and if each factor group G_i/G_{i+1} is abelian (resp. cyclic).

Let $f: G \to G'$ be a homomorphism and let

$$G' = G'_0 \supset G'_1 \supset \cdots \supset G'_m$$

be a normal tower in G'. Let $G_i = f^{-1}(G'_i)$. Then the G_i ($i = 0, \dots, m$) form a normal tower. If the G'_i form an abelian tower (resp. cyclic tower) then the G_i form an abelian tower (resp. cyclic tower), because we have an injective homomorphism

$$G_i/G_{i+1} \to G'_i/G'_{i+1}$$

for each i, and because a subgroup of an abelian group (resp. a cyclic group) is abelian (resp. cyclic).

A **refinement** of a tower

$$G = G_0 \supset G_1 \supset \cdots \supset G_m$$

is a tower which can be obtained by inserting a finite number of subgroups in the given tower. A group is said to be **solvable** if it has an abelian tower, whose last element is the trivial subgroup (i.e. $G_m = \{e\}$ in the above notation).

Proposition 3.1. *Let G be a finite group. An abelian tower of G admits a cyclic refinement. Let G be a finite solvable group. Then G admits a cyclic tower, whose last element is $\{e\}$.*

Proof. The second assertion is an immediate consequence of the first, and it clearly suffices to prove that if G is finite, abelian, then G admits a cyclic tower. We use induction on the order of G. Let x be an element of G. We may assume that $x \neq e$. Let X be the cyclic group generated by x. Let $G' = G/X$. By induction, we can find a cyclic tower in G', and its inverse image is a cyclic tower in G whose last element is X. If we refine this tower by inserting $\{e\}$ at the end, we obtain the desired cyclic tower.

Example. In Theorem 6.4 it will be proved that a group whose order is a prime power is solvable.

Example. One of the major results of group theory is the Feit-Thompson theorem that all finite groups of odd order are solvable. Cf. [Go 68].

Example. Solvable groups will occur in field theory as the Galois groups of solvable extensions. See Chapter VI, Theorem 7.2.

Example. We assume the reader knows the basic notions of linear algebra. Let k be a field. Let $G = GL(n, k)$ be the group of invertible $n \times n$ matrices in k. Let $T = T(n, k)$ be the upper triangular group; that is, the subgroup of matrices which are 0 below the diagonal. Let D be the diagonal group of diagonal matrices with non-zero components on the diagonal. Let N be the additive group of matrices which are 0 on and below the diagonal, and let $U = I + N$, where I is the unit $n \times n$ matrix. Then U is a subgroup of G. (Note that N consists of nilpotent matrices, i.e. matrices A such that $A^m = 0$ for some positive integer m. Then $(I - A)^{-1} = I + A + A^2 + \dots + A^{m-1}$ is computed using the geometric series.) Given a matrix $A \in T$, let diag(A) be the diagonal matrix which has the same diagonal components as A. Then the reader will verify that we get a surjective homomorphism

$$T \to D \quad \text{given by} \quad A \mapsto \text{diag}(A).$$

The kernel of this homomorphism is precisely U. More generally, observe that for $r \geqq 2$, the set N^{r-1} consists of all matrices of the form

$$M = \begin{pmatrix} 0 & 0 & \cdots & 0 & a_{1r} & & \cdots\cdots & a_{1n} \\ 0 & 0 & \cdots & 0 & 0 & a_{2,r+1} & \cdots & a_{2n} \\ \vdots & \vdots & & & & & \ddots & \vdots \\ 0 & 0 & \cdots\cdots & & & & & a_{n-r+1,n} \\ 0 & 0 & \cdots\cdots & & & & & 0 \\ & & & & \cdots & & & \\ 0 & 0 & \cdots\cdots & & & & & 0 \end{pmatrix}$$

Let $U_r = I + N^r$. Then $U_1 = U$ and $U_r \supset U_{r+1}$. Furthermore, U_{r+1} is normal in U_r, and the factor group is isomorphic to the additive group (!) k^{n-r}, under the the mapping which sends $I + M$ to the $n - r$-tuple $(a_{1r+1}, \dots, a_{n-r,n}) \in k^{n-r}$. This $n - r$-tuple could be called the r-th upper diagonal. Thus we obtain an abelian tower

$$T \supset U = U_1 \supset U_2 \supset \dots \supset U_n = \{I\}.$$

Theorem 3.2. *Let G be a group and H a normal subgroup. Then G is solvable if and only if H and G/H are solvable.*

Proof. We prove that G solvable implies that H is solvable. Let $G = G_0 \supset G_1 \supset \dots \supset G_r = \{e\}$ be a tower of groups with G_{i+1} normal in G_i and such that G_i/G_{i+1} is abelian. Let $H_i = H \cap G_i$. Then H_{i+1} is normal in H_i, and we have an embedding $H_i/H_{i+1} \to G_i/G_{i+1}$, whence H_i/H_{i+1} is abelian, whence proving that H is solvable. We leave the proofs of the other statements to the reader.

Let G be a group. A **commutator** in G is a group element of the form $xyx^{-1}y^{-1}$ with $x, y \in G$. Let G^c be the subgroup of G generated by the commutators. We call G^c the **commutator subgroup** of G. As an exercise, prove that G^c is normal in G, and that every homomorphism $f: G \to G'$ into a commutative group G' contains G^c in its kernel, and consequently factors through the factor commutator group G/G^c. Observe that G/G^c itself is commutative. Indeed, if $\bar{x}$ denotes the image of x in G/G^c, then by definition we have $\bar{x}\bar{y}\bar{x}^{-1}\bar{y}^{-1} = \bar{e}$, so $\bar{x}$ and $\bar{y}$ commute. In light of the definition of solvability, it is clear that the commutator group is at the heart of solvability and non-solvability problems.

A group G is said to be **simple** if it is non-trivial, and has no normal subgroups other than $\{e\}$ and G itself.

Examples. An abelian group is simple if and only if it is cyclic of prime order. Indeed, suppose A abelian and non-trivial. Let $a \in A$, $a \neq e$. If a generates an infinite cyclic group, then a^2 generates a proper subgroup and so A is not simple. If a has finite period, and A is simple, then $A = \langle a \rangle$. Let n be the period and suppose n not prime. Write $n = rs$ with $r, s > 1$. Then $a^r \neq e$ and a^r generates a proper subgroup, contradicting the simplicity of A, so a has prime period and A is cyclic of order p.

Examples. Using commutators, we shall give examples of simple groups in Theorem 5.5 (the alternating group), and in Theorem 9.2 of Chapter XIII ($PSL_n(F)$, a group of matrices to be defined in that chapter). Since a non-cyclic simple group is not solvable, we get thereby examples of non-solvable groups.

A major program of finite group theory is the classification of all finite simple groups. Essentially most of them (if not all) have natural representations as subgroups of linear maps of suitable vector spaces over suitable fields, in a suitably natural way. See [Go 82], [Go 86], [Sol 01] for surveys. Gaps in purported proofs have been found. As of 2001, these are still incomplete.

Next we are concerned with towers of subgroups such that the factor groups G_i/G_{i+1} are simple. The next lemma is for use in the proof of the Jordan-Hölder and Schreier theorems.

Lemma 3.3. **(Butterfly Lemma.)** (Zassenhaus) *Let U, V be subgroups of a group. Let u, v be normal subgroups of U and V, respectively. Then*

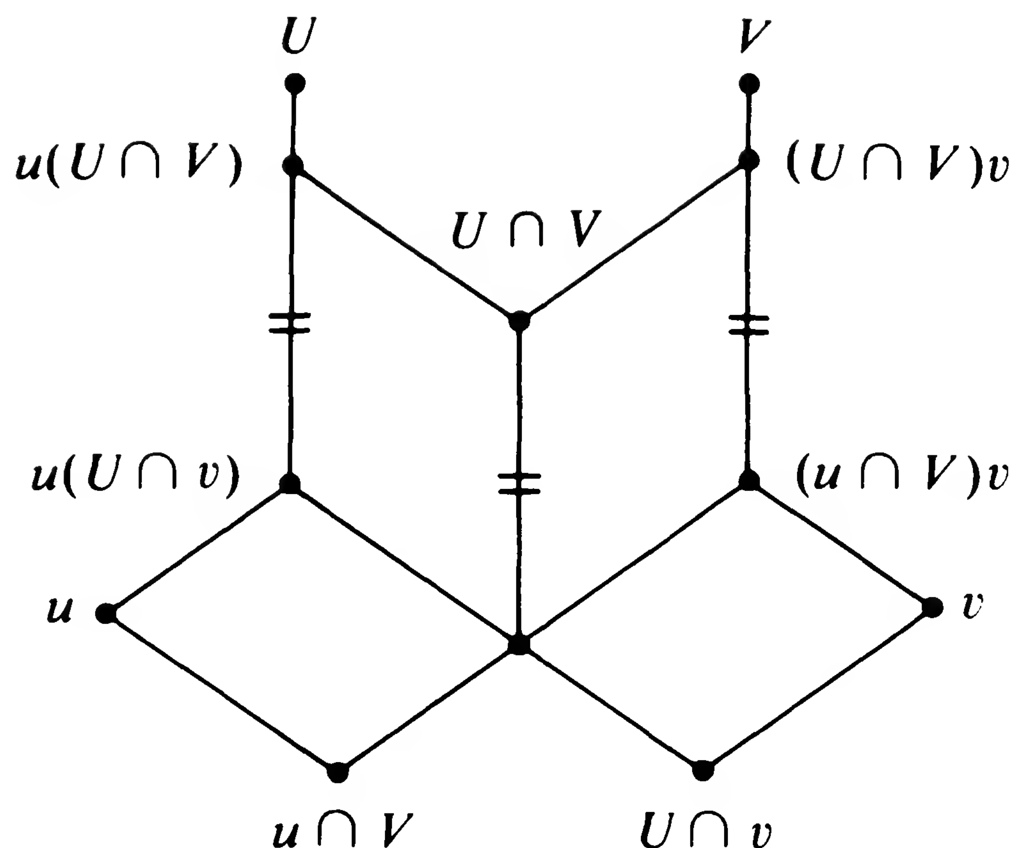
$$u(U \cap v) \quad \textit{is normal in} \quad u(U \cap V),$$

$$(u \cap V)v \quad \textit{is normal in} \quad (U \cap V)v,$$

and the factor groups are isomorphic, i.e.

$$u(U \cap V)/u(U \cap v) \approx (U \cap V)v/(u \cap V)v.$$

Proof. The combination of groups and factor groups becomes clear if one visualizes the following diagram of subgroups (which gives its name to the lemma):



In this diagram, we are given U, u, V, v. All the other points in the diagram correspond to certain groups which can be determined as follows. The intersection of two line segments going downwards represents the intersection of groups. Two lines going upwards meet in a point which represents the product of two subgroups (i.e. the smallest subgroup containing both of them).

We consider the two parallelograms representing the wings of the butterfly, and we shall give isomorphisms of the factor groups as follows:

$$\frac{u(U \cap V)}{u(U \cap v)} \approx \frac{U \cap V}{(u \cap V)(U \cap v)} \approx \frac{(U \cap V)v}{(u \cap V)v}.$$

In fact, the vertical side common to both parallelograms has $U \cap V$ as its top end point, and $(u \cap V)(U \cap v)$ as its bottom end point. We have an isomorphism

$$(U \cap V)/(u \cap V)(U \cap v) \approx u(U \cap V)/u(U \cap v).$$

This is obtained from the isomorphism theorem

$$H/(H \cap N) \approx HN/N$$

by setting $H = U \cap V$ and $N = u(U \cap v)$. This gives us the isomorphism on the left. By symmetry we obtain the corresponding isomorphism on the right, which proves the Butterfly lemma.

Let G be a group, and let

$$G = G_1 \supset G_2 \supset \cdots \supset G_r = \{e\},$$

$$G = H_1 \supset H_2 \supset \cdots \supset H_s = \{e\}$$

be normal towers of subgroups, ending with the trivial group. We shall say that these towers are **equivalent** if $r = s$ and if there exists a permutation of the

indices $i = 1, \ldots, r-1$, written $i \mapsto i'$, such that

$$G_i/G_{i+1} \approx H_{i'}/H_{i'+1}.$$

In other words, the sequences of factor groups in our two towers are the same, up to isomorphisms, and a permutation of the indices.

Theorem 3.4. (Schreier) *Let G be a group. Two normal towers of subgroups ending with the trivial group have equivalent refinements.*

Proof. Let the two towers be as above. For each $i = 1, \ldots, r-1$ and $j = 1, \ldots, s$ we define

$$G_{ij} = G_{i+1}(H_j \cap G_i).$$

Then $G_{is} = G_{i+1}$, and we have a refinement of the first tower:

$$\begin{aligned} G = G_{11} &\supset G_{12} \supset \cdots \supset G_{1,s-1} \supset G_2 \\ &= G_{21} \supset G_{22} \supset \cdots \supset G_{r-1,1} \supset \cdots \supset G_{r-1,s-1} \supset \{e\}. \end{aligned}$$

Similarly, we define

$$H_{ji} = H_{j+1}(G_i \cap H_j),$$

for $j = 1, \ldots, s-1$ and $i = 1, \ldots, r$. This yields a refinement of the second tower. By the butterfly lemma, for $i = 1, \ldots, r-1$ and $j = 1, \ldots, s-1$ we have isomorphisms

$$G_{ij}/G_{i,j+1} \approx H_{ji}/H_{j,i+1}.$$

We view each one of our refined towers as having $(r-1)(s-1)+1$ elements, namely G_{ij} $(i = 1, \ldots, r-1; j = 1, \ldots, s-1)$ and $\{e\}$ in the first case, H_{ji} and $\{e\}$ in the second case. The preceding isomorphism for each pair of indices (i, j) shows that our refined towers are equivalent, as was to be proved.

A group G is said to be **simple** if it is non-trivial, and has no normal subgroups other than $\{e\}$ and G itself.

Theorem 3.5. (Jordan-Hölder) *Let G be a group, and let*

$$G = G_1 \supset G_2 \supset \cdots \supset G_r = \{e\}$$

be a normal tower such that each group G_i/G_{i+1} is simple, and $G_i \neq G_{i+1}$ for $i = 1, \ldots, r-1$. Then any other normal tower of G having the same properties is equivalent to this one.

Proof. Given any refinement $\{G_{ij}\}$ as before for our tower, we observe that for each i, there exists precisely one index j such that $G_i/G_{i+1} = G_{ij}/G_{i,j+1}$. Thus the sequence of non-trivial factors for the original tower, or the refined tower, is the same. This proves our theorem.

Bibliography

[Go 68] D. Gorenstein, *Finite groups*, Harper and Row, 1968

[Go 82] D. Gorenstein, *Finite simple groups*, Plenum Press, 1982

[Go 83] D. Gorenstein, *The Classification of Finite Simple Groups*, Plenum Press, 1983

[Go 86] D. Gorenstein, Classifying the finite simple groups, *Bull. AMS* **14** No. 1 (1986), pp. 1–98

[So 01] R. Solomon, A brief history of the classification of the finite simple groups, *Bull. AMS* **38**, 3 (2001) pp. 315–352

§4. CYCLIC GROUPS

The integers $\mathbf{Z}$ form an additive group. We shall determine its subgroups. Let H be a subgroup of $\mathbf{Z}$. If H is not trivial, let a be the smallest positive integer in H. We contend that H consists of all elements na, with $n \in \mathbf{Z}$. To prove this, let $y \in H$. There exist integers n, r with $0 \leqq r < a$ such that

$$y = na + r.$$

Since H is a subgroup and $r = y - na$, we have $r \in H$, whence $r = 0$, and our assertion follows.

Let G be a group. We shall say that G is **cyclic** if there exists an element a of G such that every element x of G can be written in the form a^n for some $n \in \mathbf{Z}$ (in other words, if the map $f: \mathbf{Z} \to G$ such that $f(n) = a^n$ is surjective). Such an element a of G is then called a **generator** of G.

Let G be a group and $a \in G$. The subset of all elements a^n $(n \in \mathbf{Z})$ is obviously a subgroup of G, which is cyclic. If m is an integer such that $a^m = e$ and $m > 0$ then we shall call m an **exponent** of a. We shall say that $m > 0$ is an **exponent** of G if $x^m = e$ for all $x \in G$.

Let G be a group and $a \in G$. Let $f: \mathbf{Z} \to G$ be the homomorphism such that $f(n) = a^n$ and let H be the kernel of f. Two cases arise:

1. The kernel is trivial. Then f is an isomorphism of $\mathbf{Z}$ onto the cyclic subgroup of G generated by a, and this subgroup is infinite cyclic. If a generates G, then G is cyclic. We also say that a has **infinite period**.

2. The kernel is not trivial. Let d be the smallest positive integer in the kernel. Then d is called the **period** of a. If m is an integer such that $a^m = e$ then $m = ds$ for some integer s. We observe that the elements $e, a, \dots, a^{d-1}$ are

distinct. Indeed, if $a^r = a^s$ with $0 \leqq r, s \leqq d - 1$, and say $r \leqq s$, then $a^{s-r} = e$. Since $0 \leqq s - r < d$ we must have $s - r = 0$. The cyclic subgroup generated by a has order d. Hence by Proposition 2.2:

Proposition 4.1. *Let G be a finite group of order $n > 1$. Let a be an element of G, $a \neq e$. Then the period of a divides n. If the order of G is a prime number p, then G is cyclic and the period of any generator is equal to p.*

Furthermore:

Proposition 4.2. *Let G be a cyclic group. Then every subgroup of G is cyclic. If f is a homomorphism of G, then the image of f is cyclic.*

Proof. If G is infinite cyclic, it is isomorphic to $\mathbf{Z}$, and we determined above all subgroups of $\mathbf{Z}$, finding that they are all cyclic. If $f : G \to G'$ is a homomorphism, and a is a generator of G, then $f(a)$ is obviously a generator of $f(G)$, which is therefore cyclic, so the image of f is cyclic. Next let H be a subgroup of G. We want to show H cyclic. Let a be a generator of G. Then we have a surjective homomorphism $f : \mathbf{Z} \to G$ such that $f(n) = a^n$. The inverse image $f^{-1}(H)$ is a subgroup of $\mathbf{Z}$, and therefore equal to $m\mathbf{Z}$ for some positive integer m. Since f is surjective, we also have a surjective homomorphism $m\mathbf{Z} \to H$. Since $m\mathbf{Z}$ is cyclic (generated additively by m), it follows that H is cyclic, thus proving the proposition.

We observe that two cyclic groups of the same order m are isomorphic. Indeed, if G is cyclic of order m with generator a, then we have a surjective homomorphism $f: \mathbf{Z} \to G$ such that $f(n) = a^n$, and if $k\mathbf{Z}$ is the kernel, with k positive, then we have an isomorphism $\mathbf{Z}/k\mathbf{Z} \approx G$, so $k = m$. If $u: G_1 \to \mathbf{Z}/m\mathbf{Z}$ and $v: G_2 \to \mathbf{Z}/m\mathbf{Z}$ are isomorphisms of two cyclic groups with $\mathbf{Z}/m\mathbf{Z}$, then $v^{-1} \circ u: G_1 \to G_2$ is an isomorphism.

Proposition 4.3.

(i) *An infinite cyclic group has exactly two generators (if a is a generator, then a^{-1} is the only other generator).*

(ii) *Let G be a finite cyclic group of order n, and let x be a generator. The set of generators of G consists of those powers x^ν of x such that ν is relatively prime to n.*

(iii) *Let G be a cyclic group, and let a, b be two generators. Then there exists an automorphism of G mapping a onto b. Conversely, any automorphism of G maps a on some generator of G.*

(iv) *Let G be a cyclic group of order n. Let d be a positive integer dividing n. Then there exists a unique subgroup of G of order d.*

(v) *Let G_1, G_2 be cyclic of orders m, n respectively. If m, n are relatively prime then $G_1 \times G_2$ is cyclic.*

(vi) *Let G be a finite abelian group. If G is not cyclic, then there exists a prime p and a subgroup of G isomorphic to $C \times C$, where C is cyclic of order p.*

Proof. We leave the first three statements to the reader, and prove the others.

(iv) Let $d|n$. Let $m = n/d$. Let $f : \mathbf{Z} \to G$ be a surjective homomorphism. Then $f(m\mathbf{Z})$ is a subgroup of G, and from the isomorphism $\mathbf{Z}/m\mathbf{Z} \approx G/f(m\mathbf{Z})$ we conclude that $f(m\mathbf{Z})$ has index m in G, whence $f(m\mathbf{Z})$ has order d. Conversely, let H be a subgroup of order d. Then $f^{-1}(H) = m\mathbf{Z}$ for some positive integer m, so $H = f(m\mathbf{Z})$, $\mathbf{Z}/m\mathbf{Z} \approx G/H$, so $n = md$, $m = n/d$ and H is uniquely determined.

(v) Let $A = \langle a \rangle$ and $B = \langle b \rangle$ be cyclic groups of orders m, n, relatively prime. Consider the homomorphism $\mathbf{Z} \to A \times B$ such that $k \mapsto (a^k, b^k)$. An element in its kernel must be divisible both by m and n, hence by their product since m, n are relatively prime. Conversely, it is clear that $mn\mathbf{Z}$ is contained in the kernel, so the kernel is $mn\mathbf{Z}$. The image of $\mathbf{Z} \to A \times B$ is surjective by the Chinese remainder theorem. This proves (v). (A reader who does not know the Chinese remainder theorem can see a proof in the more general context of Chapter II, Theorem 2.2.)

(vi) This characterization of cyclic groups is an immediate consequence of the structure theorem which will be proved in §8, because if G is not cyclic, then by Theorem 8.1 and (v) we are reduced to the case when G is a p-group, and by Theorem 8.2 there are at least two factors in the direct product (or sum) decomposition, and each contains a cyclic subgroup of order p, whence G contains their direct product (or sum). Statement (vi) is, of course, easier to prove than the full structure theorem, and it is a good exercise for the reader to formulate the simpler arguments which yield (vi) directly.

Note. For the group of automorphisms of a cyclic group, see the end of Chapter II, §2.

§5. OPERATIONS OF A GROUP ON A SET

Let G be a group and let S be a set. An **operation** or an **action** of G on S is a homomorphism

$$\pi : G \to \mathrm{Perm}(S)$$

of G into the group of permutations of S. We then call S a **G-set**. We denote the permutation associated with an element $x \in G$ by π_x. Thus the homomorphism is denoted by $x \mapsto \pi_x$. Given $s \in S$, the image of s under the permutation π_x is $\pi_x(s)$. From such an operation we obtain a mapping

$$G \times S \to S,$$

which to each pair (x, s) with $x \in G$ and $s \in S$ associates the element $\pi_x(s)$. We often abbreviate the notation and write simply xs instead of $\pi_x(s)$. With the simpler notation, we have the two properties:

For all $x, y \in G$ and $s \in S$, we have $x(ys) = (xy)s$.
If e is the unit element of G, then $es = s$ for all $s \in S$.

Conversely, if we are given a mapping $G \times S \to S$, denoted by $(x, s) \mapsto xs$, satisfying these two properties, then for each $x \in G$ the map $s \mapsto xs$ is permutation of S, which we then denote by $\pi_x(s)$. Then $x \mapsto \pi_x$ is a homomorphism of G into Perm(S). So an operation of G on S could also be defined as a mapping $G \times S \to S$ satisfying the above two properties. The most important examples of representations of G as a group of permutations are the following.

1. Conjugation. For each $x \in G$, let $\mathbf{c}_x: G \to G$ be the map such that $\mathbf{c}_x(y) = xyx^{-1}$. Then it is immediately verified that the association $x \mapsto \mathbf{c}_x$ is a homomorphism $G \to \mathrm{Aut}(G)$, and so this map gives an operation of G on itself, called **conjugation**. The kernel of the homomorphism $x \mapsto \mathbf{c}_x$ is a normal subgroup of G, which consists of all $x \in G$ such that $xyx^{-1} = y$ for all $y \in G$, i.e. all $x \in G$ which commute with every element of G. This kernel is called the **center** of G. Automorphisms of G of the form $\mathbf{c}_x$ are called **inner**.

To avoid confusion about the operation on the left, we don't write xy for $\mathbf{c}_x(y)$. Sometimes, one writes

$$\mathbf{c}_{x^{-1}}(y) = x^{-1}yx = y^x,$$

i.e. one uses an exponential notation, so that we have the rules

$$y^{(xz)} = (y^x)^z \quad \text{and} \quad y^e = y$$

for all $x, y, z \in G$. Similarly, ${}^xy = xyx^{-1}$ and ${}^z({}^xy) = {}^{zx}y$.

We note that G also operates by conjugation on the set of subsets of G. Indeed, let S be the set of subsets of G, and let $A \in S$ be a subset of G. Then xAx^{-1} is also a subset of G which may be denoted by $\mathbf{c}_x(A)$, and one verifies trivially that the map

$$(x, A) \mapsto xAx^{-1}$$

of $G \times S \to S$ is an operation of G on S. We note in addition that if A is a subgroup of G then xAx^{-1} is also a subgroup, so that G operates on the set of subgroups by conjugation.

If A, B are two subsets of G, we say that they are **conjugate** if there exists $x \in G$ such that $B = xAx^{-1}$.

2. Translation. For each $x \in G$ we define the translation $T_x: G \to G$ by $T_x(y) = xy$. Then the map

$$(x, y) \mapsto xy = T_x(y)$$

defines an operation of G on itself. *Warning*: T_x is not a group-homomorphism! Only a permutation of G.

Similarly, G operates by translation on the set of subsets, for if A is a subset of G, then $xA = T_x(A)$ is also a subset. If H is a subgroup of G, then $T_x(H) = xH$ is in general not a subgroup but a coset of H, and hence we see that G operates by translation on the set of cosets of H. We denote the set of left cosets of H by G/H. Thus even though H need not be normal, G/H is a G-set. It has become customary to denote the set of *right* cosets by $H\backslash G$.

The above two representations of G as a group of permutations will be used frequently in the sequel. In particular, the representation by conjugation will be used throughout the next section, in the proof of the Sylow theorems.

3. Example from linear algebra. We assume the reader knows basic notions of linear algebra. Let k be a field and let V be a vector space over k. Let $G = GL(V)$ be the group of linear automorphisms of V. For $A \in G$ and $v \in V$, the map $(A, v) \mapsto Av$ defines an operation of G on V. Of course, G is a subgroup of the group of permutations $\mathrm{Perm}(V)$. Similarly, let $V = k^n$ be the vector space of (vertical) n-tuples of elements of k, and let G be the group of invertible $n \times n$ matrices with components in k. Then G operates on k^n by $(A, X) \mapsto AX$ for $A \in G$ and $X \in k^n$.

Let S, S' be two G-sets, and $f : S \to S'$ a map. We say that f is a **morphism of G-sets**, or a **G-map**, if

$$f(xs) = xf(s)$$

for all $x \in G$ and $s \in S$. (We shall soon define categories, and see that G-sets form a category.)

We now return to the general situation, and consider a group operating on a set S. Let $s \in S$. The set of elements $x \in G$ such that $xs = s$ is obviously a subgroup of G, called the **isotropy** group of s in G, and denoted by G_s.

When G operates on itself by conjugation, then the isotropy group of an element is none other than the normalizer of this element. Similarly, when G operates on the set of subgroups by conjugation, the isotropy group of a subgroup is again its normalizer.

Let G operate on a set S. Let s, s' be elements of S, and y an element of G such that $ys = s'$. Then

$$G_{s'} = yG_sy^{-1}$$

Indeed, one sees at once that yG_sy^{-1} leaves s' fixed. Conversely, if $x's' = s'$ then $x'ys = ys$, so $y^{-1}x'y \in G_s$ and $x' \in yG_sy^{-1}$. Thus the isotropy groups of s and s' are conjugate.

Let K be the kernel of the representation $G \to \mathrm{Perm}(S)$. Then directly from the definitions, we obtain that

$$K = \bigcap_{s \in S} G_s = \text{intersection of all isotropy groups.}$$

An action or operation of G is said to be **faithful** if $K = \{e\}$; that is, the kernel of $G \to \text{Perm}(S)$ is trivial. A **fixed point** of G is an element $s \in S$ such that $xs = s$ for all $x \in G$ or in other words, $G = G_s$.

Let G operate on a set S. Let $s \in S$. The subset of S consisting of all elements xs (with $x \in G$) is denoted by Gs, and is called the **orbit** of s under G. If x and y are in the same coset of the subgroup $H = G_s$, then $xs = ys$, and conversely (obvious). In this manner, we get a mapping

$$f: G/H \to S$$

given by $f(xH) = xs$, and it is clear that this map is a morphism of G-sets. In fact, one sees at once that it induces a bijection of G/H onto the orbit Gs. Consequently:

Proposition 5.1. *If G is a group operating on a set S, and $s \in S$, then the order of the orbit Gs is equal to the index $(G : G_s)$.*

In particular, when G operates by conjugation on the set of subgroups, and H is a subgroup, then:

Proposition 5.2. *The number of conjugate subgroups to H is equal to the index of the normalizer of H.*

Example. Let G be a group and H a subgroup of index 2. Then H is normal in G.

Proof. Note that H is contained in its normalizer N_H, so the index of N_H in G is 1 or 2. If it is 1, then we are done. Suppose it is 2. Let G operate by conjugation on the set of subgroups. The orbit of H has 2 elements, and G operates on this orbit. In this way we get a homomorphism of G into the group of permutations of 2 elements. Since there is one conjugate of H unequal to H, then the kernel of our homomorphism is normal, of index 2, hence equal to H, which is normal, a contradiction which concludes the proof.

For a generalization and other examples, see Lemma 6.7.

In general, an operation of G on S is said to be **transitive** if there is only one orbit.

Examples. The symmetric group S_n operates transitively on $\{1, 2, \ldots, n\}$. In Proposition 2.1 of Chapter VII, we shall see a non-trivial example of transitive action of a Galois group operating on the primes lying above a given prime in the ground ring. In topology, suppose we have a universal covering space $p: X' \to X$, where X is connected. Given $x \in X$, the fundamental group $\pi_1(X)$ operates transitively on the inverse image $p^{-1}(x)$.

Example. Let $\mathfrak{H}$ be the upper half-plane; that is, the set of complex numbers $z = x + iy$ such that $y > 0$. Let $G = SL_2(\mathbf{R})$ (2×2 matrices with determinant 1). For

$$\alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G, \text{ we let } \alpha z = \frac{az + b}{cz + d}.$$

Readers will verify by brute force that this defines an operation of G on $\mathfrak{H}$. The isotropy group of i is the group of matrices

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \quad \text{with } \theta \text{ real.}$$

This group is usually denoted by K. The group G operates transitively. You can verify all these statements as easy exercises.

Let G operate on a set S. Then two orbits of G are either disjoint or are equal. Indeed, if Gs_1 and Gs_2 are two orbits with an element s in common, then $s = xs_1$ for some $x \in G$, and hence $Gs = Gxs_1 = Gs_1$. Similarly, $Gs = Gs_2$. Hence S is the disjoint union of the distinct orbits, and we can write

$$S = \bigcup_{i \in I} Gs_i \qquad \text{(disjoint)}, \qquad \text{also denoted } S = \coprod_{i \in I} Gs_i,$$

where I is some indexing set, and the s_i are elements of distinct orbits. If S is finite, this gives a decomposition of the order of S as a sum of orders of orbits, which we call the **orbit decomposition formula**, namely

$$\boxed{\operatorname{card}(S) = \sum_{i \in I} (G : G_{s_i}).}$$

Let x, y be elements of a group (or monoid) G. They are said to **commute** if $xy = yx$. If G is a group, the set of all elements $x \in G$ which commute with all elements of G is a subgroup of G which we called the **center** of G. Let G act on itself by conjugation. Then x is in the center if and only if the orbit of x is x itself, and thus has one element. In general, the order of the orbit of x is equal to the index of the normalizer of x. Thus when G is a finite group, the above formula reads

$$\boxed{(G : 1) = \sum_{x \in C} (G : G_x)}$$

where C is a set of representatives for the distinct conjugacy classes, and the sum is taken over all $x \in C$. This formula is also called the **class formula**.

The class formula and the orbit decomposition formula will be used systematically in the next section on Sylow groups, which may be viewed as providing examples for these formulas.

Readers interested in Sylow groups may jump immediately to the next section. The rest of this section deals with special properties of the symmetric group, which may serve as examples of the general notions we have developed.

The symmetric group. Let S_n be the group of permutations of a set with n elements. This set may be taken to be the set of integers $J_n = \{1, 2, \dots, n\}$. Given any $\sigma \in S_n$, and any integer i, $1 \leqq i \leqq n$, we may form the orbit of i under the cyclic group generated by σ. Such an orbit is called a **cycle** for σ, and may be written

$$[i_1 i_2 \cdots i_r], \qquad \text{so} \quad \sigma(i_1) = i_2, \dots, \sigma(i_{r-1}) = i_r, \sigma(i_r) = i_1.$$

Then $\{1, \dots, n\}$ may be decomposed into a disjoint union of orbits for the cyclic group generated by σ, and therefore into disjoint cycles. Thus the effect of σ on $\{1, \dots, n\}$ is represented by a product of disjoint cycles.

Example. The cycle [132] represents the permutation σ such that

$$\sigma(1) = 3, \qquad \sigma(3) = 2, \quad \text{and} \quad \sigma(2) = 1.$$

We have $\sigma^2(1) = 2$, $\sigma^3(1) = 1$. Thus $\{1, 3, 2\}$ is the orbit of 1 under the cyclic group generated by σ.

Example. In Exercise 38, one will see how to generate S_n by special types of generators. Perhaps the most important part of that exercise is that if n is prime, σ is an n-cycle and τ is a transposition, then σ, τ generate S_n. As an application in Galois theory, if one tries to prove that a Galois group is all of S_n (as a group of permutations of the roots), it suffices to prove that the Galois group contains an n-cycle and a transposition. See Example 6 of Chapter VI, §2.

We want to associate a sign ± 1 to each permutation. We do this in the standard way. Let f be a function of n variables, say $f : \mathbf{Z}^n \to \mathbf{Z}$, so we can evaluate $f(x_1, \dots, x_n)$. Let σ be a permutation of J_n. We define the function $\pi(\sigma)f$ by

$$\pi(\sigma)f(x_1, \dots, x_n) = f(x_{\sigma(1)}, \dots, x_{\sigma(n)}).$$

Then for $\sigma, \tau \in S_n$ we have $\pi(\sigma\tau) = \pi(\sigma)\pi(\tau)$. Indeed, we use the definition applied to the function $g = \pi(\tau)f$ to get

$$\begin{aligned}\pi(\sigma)\pi(\tau)f(x_1, \dots, x_n) &= (\pi(\tau)f)(x_{\sigma(1)}, \dots, x_{\sigma(n)}) \\ &= f(x_{\sigma\tau(1)}, \dots, x_{\sigma\tau(n)}) \\ &= \pi(\sigma\tau)f(x_1, \dots, x_n).\end{aligned}$$

Since the identity in S_n operates as the identity on functions, it follows that we have obtained an operation of S_n on the set of functions. We shall write more simply σf instead of $\pi(\sigma)f$. It is immediately verified that for two functions f, g we have

$$\sigma(f + g) = \sigma f + \sigma g \quad \text{and} \quad \sigma(fg) = (\sigma f)(\sigma g).$$

If c is constant, then $\sigma(cf) = c\sigma(f)$.

Proposition 5.3. *There exists a unique homomorphism $\varepsilon: S_n \to \{\pm 1\}$ such that for every transposition τ we have $\varepsilon(\tau) = -1$.*

Proof. Let Δ be the function

$$\Delta(x_1, \ldots, x_n) = \prod_{i<j} (x_j - x_i),$$

the product being taken for all pairs of integers i, j satisfying $1 \leqq i < j \leqq n$. Let τ be a transposition, interchanging the two integers r and s. Say $r < s$. We wish to determine

$$\tau\Delta(x_1, \ldots, x_n) = \prod_{i<j} (x_{\tau(j)} - x_{\tau(i)}).$$

For one factor involving $j = s$, $i = r$, we see that τ changes the factor $(x_s - x_r)$ to $-(x_s - x_r)$. All other factors can be considered in pairs as follows:

$$\begin{aligned}
&(x_k - x_s)(x_k - x_r) \quad \text{if } k > s, \\
&(x_s - x_k)(x_k - x_r) \quad \text{if } r < k < s, \\
&(x_s - x_k)(x_r - x_k) \quad \text{if } k < r.
\end{aligned}$$

Each one of these pairs remains unchanged when we apply τ. Hence we see that $\tau\Delta = -\Delta$.

Let $\varepsilon(\sigma)$ be the sign 1 or -1 such that $\sigma\Delta = \varepsilon(\sigma)\Delta$ for a permutation σ. Since $\pi(\sigma\tau) = \pi(\sigma)\pi(\tau)$, it follows at once that ε is a homomorphism, and the proposition is proved.

In particular, if $\sigma = \tau_1 \cdots \tau_m$ is a product of transpositions, then $\varepsilon(\sigma) = (-1)^m$. As a matter of terminology, we call σ **even** if $\varepsilon(\sigma) = 1$, and **odd** if $\varepsilon(\sigma) = -1$. The even permutations constitute the kernel of ε, which is called the **alternating group** A_n.

Theorem 5.4. *If $n \geqq 5$ then S_n is not solvable.*

Proof. We shall first prove that if H, N are two subgroups of S_n such that $N \subset H$ and N is normal in H, if H contains every 3-cycle, and if H/N is abelian, then N contains every 3-cycle. To see this, let i, j, k, r, s be five distinct integers in J_n, and let $\sigma = [ijk]$ and $\tau = [krs]$. Then a direct computation gives their commutator

$$\sigma\tau\sigma^{-1}\tau^{-1} = [rki].$$

Since the choice of i, j, k, r, s was arbitrary, we see that the cycles $[rki]$ all lie in N for all choices of distinct r, k, i, thereby proving what we wanted.

Now suppose that we have a tower of subgroups

$$S_n = H_0 \supset H_1 \supset H_2 \supset \cdots \supset H_m = \{e\}$$

such that H_v is normal in H_{v-1} for $v = 1, \ldots, m$, and H_v/H_{v-1} is abelian. Since S_n contains every 3-cycle, we conclude that H_1 contains every 3-cycle. By induction, we conclude that $H_m = \{e\}$ contains every 3-cycle, which is impossible, thus proving the theorem.

Remark concerning the sign $\varepsilon(\sigma)$. *A priori*, we defined the sign for a given n, so we should write $\varepsilon_n(\sigma)$. However, suppose $n < m$. Then the restriction of ε_m to S_n (viewed as a permutation of J_n leaving the elements of J_m not in J_n fixed) gives a homomorphism satisfying the conditions of Proposition 5.3, so this restriction is equal to ε_n. Thus $A_m \cap S_n = A_n$.

Next we prove some properties of the alternating group.

(a) *A_n is generated by the 3-cycles. Proof:* Consider the product of two transpositions $[ij][rs]$. If they have an element in common, the product is either the identity or a 3-cycle. If they have no element in common, then

$$[ij][rs] = [ijr][jrs],$$

so the product of two transpositions is also a product of 3-cycles. Since an even permutation is a product of an even number of transpositions, we are done.

(b) *If $n \geqq 5$, all 3-cycles are conjugate in A_n. Proof:* If γ is a permutation, then for a cycle $[i_1 \ldots i_m]$ we have

$$\gamma[i_1 \ldots i_m]\gamma^{-1} = [\gamma(i_1) \ldots \gamma(i_m)].$$

Given 3-cycles $[ijk]$ and $[i'j'k']$ there is a permutation γ such that $\gamma(i) = i'$, $\gamma(j) = j'$, and $\gamma(k) = k'$. Thus two 3-cycles are conjugate in S_n by some element γ. If γ is even, we are done. Otherwise, by assumption $n \geqq 5$ there exist r, s not equal to any one of the three elements i, j, k. Then $[rs]$ commutes with $[ijk]$, and we replace γ by $\gamma[rs]$ to prove (b).

Theorem 5.5. *If $n \geqq 5$ then the alternating group A_n is simple.*

Proof. Let N be a non-trivial normal subgroup of A_n. We prove that N contains some 3-cycle, whence the theorem follows by (b). Let $\sigma \in N$, $\sigma \neq id$, be an element which has the maximal number of fixed points; that is, integers i such that $\sigma(i) = i$. It will suffice to prove that σ is a 3-cycle or the identity. Decompose J_n into disjoint orbits of $\langle\sigma\rangle$. Then some orbits have more than one element. Suppose all orbits have 2 elements (except for the fixed points). Since σ is even, there are at least two such orbits. On their union, σ is represented as

a product of two transpositions $[ij][rs]$. Let $k \neq i, j, r, s$. Let $\tau = [rsk]$. Let $\sigma' = \tau\sigma\tau^{-1}\sigma^{-1}$. Then σ' is a product of a conjugate of σ and σ^{-1}, so $\sigma' \in N$. But σ' leaves i, j fixed, and any element $t \in J_n$, $t \neq i, j, r, s, k$ left fixed by σ is also fixed by σ', so σ' has more fixed points than σ, contradicting our hypothesis.

So we are reduced to the case when at least one orbit of $\langle\sigma\rangle$ has $\geqq 3$ elements, say $i, j, k, \ldots$. If σ is not the 3-cycle $[ijk]$, then σ must move at least two other elements of J_n, otherwise σ is an odd permutation $[ijkr]$ for some $r \in J_n$, which is impossible. Then let σ move r, s other than i, j, k, and let $\tau = [krs]$. Let σ' be the commutator as before. Then $\sigma' \in N$ and $\sigma'(i) = i$, and all fixed points of σ are also fixed points of σ' whence σ' has more fixed points than σ, a contradiction which proves the theorem.

Example. For $n = 4$, the group A_4 is not simple. As an exercise, show that A_4 contains a unique subgroup of order 4, which is not cyclic, and which is normal. This subgroup is also normal in S_4. Write down explicitly its elements as products of transpositions.

§6. SYLOW SUBGROUPS

Let p be a prime number. By a ***p*-group**, we mean a finite group whose order is a power of p (i.e. p^n for some integer $n \geqq 0$). Let G be a finite group and H a subgroup. We call H a ***p*-subgroup** of G if H is a p-group. We call H a ***p*-Sylow** subgroup if the order of H is p^n and if p^n is the highest power of p dividing the order of G. We shall prove below that such subgroups always exist. For this we need a lemma.

Lemma 6.1. *Let G be a finite abelian group of order m, let p be a prime number dividing m. Then G has a subgroup of order p.*

Proof. We first prove by induction that if G has exponent n then the order of G divides some power of n. Let $b \in G$, $b \neq 1$, and let H be the cyclic subgroup generated by b. Then the order of H divides n since $b^n = 1$, and n is an exponent for G/H. Hence the order of G/H divides a power of n by induction, and consequently so does the order of G because

$$(G:1) = (G:H)(H:1).$$

Let G have order divisible by p. By what we have just seen, there exists an element x in G whose period is divisible by p. Let this period be ps for some integer s. Then $x^s \neq 1$ and obviously x^s has period p, and generates a subgroup of order p, as was to be shown.

Theorem 6.2. *Let G be a finite group and p a prime number dividing the order of G. Then there exists a p-Sylow subgroup of G.*

Proof. By induction on the order of G. If the order of G is prime, our assertion is obvious. We now assume given a finite group G, and assume the theorem proved for all groups of order smaller than that of G. If there exists a proper subgroup H of G whose index is prime to p, then a p-Sylow subgroup of H will also be one of G, and our assertion follows by induction. We may therefore assume that every proper subgroup has an index divisible by p. We now let G act on itself by conjugation. From the class formula we obtain

$$(G:1) = (Z:1) + \sum (G:G_x).$$

Here, Z is the center of G, and the term $(Z:1)$ corresponds to the orbits having one element, namely the elements of Z. The sum on the right is taken over the other orbits, and each index $(G:G_x)$ is then > 1, hence divisible by p. Since p divides the order of G, it follows that p divides the order of Z, hence in particular that G has a non-trivial center.

Let a be an element of order p in Z, and let H be the cyclic group generated by a. Since H is contained in Z, it is normal. Let $f: G \to G/H$ be the canonical map. Let p^n be the highest power of p dividing $(G:1)$. Then p^{n-1} divides the order of G/H. Let K' be a p-Sylow subgroup of G/H (by induction) and let $K = f^{-1}(K')$. Then $K \supset H$ and f maps K onto K'. Hence we have an isomorphism $K/H \approx K'$. Hence K has order $p^{n-1}p = p^n$, as desired.

For the rest of the theorems, we systematically use the notion of a fixed point. Let G be a group operating on a set S. Recall that a **fixed point** s of G in S is an element s of S such that $xs = s$ for all $x \in G$.

Lemma 6.3. *Let H be a p-group acting on a finite set S. Then:*

(a) *The number of fixed points of H is $\equiv \#(S) \bmod p$.*

(b) *If H has exactly one fixed point, then $\#(S) \equiv 1 \bmod p$.*

(c) *If $p \mid \#(S)$, then the number of fixed points of H is $\equiv 0 \bmod p$.*

Proof. We repeatedly use the orbit formula

$$\#(S) = \sum (H : H_{s_i}).$$

For each fixed point s_i we have $H_{s_i} = H$. For s_i not fixed, the index $(H : H_{s_i})$ is divisible by p, so (a) follows at once. Parts (b) and (c) are special cases of (a), thus proving the lemma.

Remark. In Lemma 6.3(c), if H has one fixed point, then H has at least p fixed points.

Theorem 6.4. *Let G be a finite group.*

(i) *If H is a p-subgroup of G, then H is contained in some p-Sylow subgroup.*

(ii) *All p-Sylow subgroups are conjugate.*

(iii) *The number of p-Sylow subgroups of G is* $\equiv 1 \bmod p$.

Proof. Let P be a p-Sylow subgroup of G. Suppose first that H is contained in the normalizer of P. We prove that $H \subset P$. Indeed, HP is then a subgroup of the normalizer, and P is normal in HP. But

$$(HP : P) = (H : H \cap P),$$

so if $HP \neq P$, then HP has order a power of p, and the order is larger than $\#(P)$, contradicting the hypothesis that P is a Sylow group. Hence $HP = P$ and $H \subset P$.

Next, let S be the set of all conjugates of P in G. Then G operates on S by conjugation. Since the normalizer of P contains P, and has therefore index prime to p, it follows that $\#(S)$ is not divisible by p. Now let H be any p-subgroup. Then H also acts on S by conjugation. By Lemma 6.3(a), we know that H cannot have 0 fixed points. Let Q be a fixed point. By definition this means that H is contained in the normalizer of Q, and hence by the first part of the proof, that $H \subset Q$, which proves the first part of the theorem. The second part follows immediately by taking H to be a p-Sylow group, so $\#(H) = \#(Q)$, whence $H = Q$. In particular, when H is a p-Sylow group, we see that H has only one fixed point, so that **(iii)** follows from Lemma 6.3(b). This proves the theorem.

Theorem 6.5. *Let G be a finite p-group. Then G is solvable. If its order is* > 1, *then G has a non-trivial center.*

Proof. The first assertion follows from the second, since if G has center Z, and we have an abelian tower for G/Z by induction, we can lift this abelian tower to G to show that G is solvable. To prove the second assertion, we use the class equation

$$(G:1) = \text{card}(Z) + \sum (G:G_x),$$

the sum being taken over certain x for which $(G:G_x) \neq 1$. Then p divides $(G:1)$ and also divides every term in the sum, so that p divides the order of the center, as was to be shown.

Corollary 6.6. *Let G be a p-group which is not of order* 1. *Then there exists a sequence of subgroups*

$$\{e\} = G_0 \subset G_1 \subset G_2 \subset \cdots \subset G_n = G$$

such that G_i *is normal in G and* G_{i+1}/G_i *is cyclic of order p.*

Proof. Since G has a non-trivial center, there exists an element $a \neq e$ in the center of G, and such that a has order p. Let H be the cyclic group generated by a. By induction, if $G \neq H$, we can find a sequence of subgroups as stated above in the factor group G/H. Taking the inverse image of this tower in G gives us the desired sequence in G.

We now give some examples to show how to put some of the group theory together.

Lemma 6.7. *Let G be a finite group and let p be the smallest prime dividing the order of G. Let H be a subgroup of index p. Then H is normal.*

Proof. Let $N(H) = N$ be the normalizer of H. Then $N = G$ or $N = H$. If $N = G$ we are done. Suppose $N = H$. Then the orbit of H under conjugation has $p = (G : H)$ elements, and the representation of G on this orbit gives a homomorphism of G into the symmetric group on p elements, whose order is $p!$. Let K be the kernel. Then K is the intersection of the isotropy groups, and the isotropy group of H is H by assumption, so $K \subset H$. If $K \neq H$, then from

$$(G : K) = (G : H)(H : K) = p(H : K),$$

and the fact that only the first power of p divides $p!$, we conclude that some prime dividing $(p - 1)!$ also divides $(H : K)$, which contradicts the assumption that p is the smallest prime dividing the order of G, and proves the lemma.

Proposition 6.8. *Let p, q be distinct primes and let G be a group of order pq. Then G is solvable.*

Proof. Say $p < q$. Let Q be a Sylow subgroup of order q. Then Q has index p, so by the lemma, Q is normal and the factor group has order p. But a group of prime order is cyclic, whence the proposition follows.

Example. Let G be a group of order 35. We claim that G is cyclic.

Proof. Let H_7 be the Sylow subgroup of order 7. Then H_7 is normal by Lemma 6.7. Let H_5 be a 5-Sylow subgroup, which is of order 5. Then H_5 operates by conjugation on H_7, so we get a homomorphism $H_5 \to \operatorname{Aut}(H_7)$. But $\operatorname{Aut}(H_7)$ is cyclic of order 6, so $H_5 \to \operatorname{Aut}(H_7)$ is trivial, so every element of H_5 commutes with elements of H_7. Let $H_5 = \langle x \rangle$ and $H_7 = \langle y \rangle$. Then x, y commute with each other and with themselves, so G is abelian, and so G is cyclic by Proposition 4.3(v).

Example. The techniques which have been developed are sufficient to treat many cases of the above types. For instance every group of order < 60 is solvable, as you will prove in Exercise 27.

§7. DIRECT SUMS AND FREE ABELIAN GROUPS

Let $\{A_i\}_{i \in I}$ be a family of abelian groups. We define their **direct sum**

$$A = \bigoplus_{i \in I} A_i$$

to be the subset of the direct product $\prod A_i$ consisting of all families $(x_i)_{i \in I}$ with

$x_i \in A_i$ such that $x_i = 0$ for all but a finite number of indices i. Then it is clear that A is a subgroup of the product. For each index $j \in I$, we map

$$\lambda_j : A_j \to A$$

by letting $\lambda_j(x)$ be the element whose j-th component is x, and having all other components equal to 0. Then λ_i is an injective homomorphism.

Proposition 7.1. *Let $\{f_i : A_i \to B\}$ be a family of homomorphisms into an abelian group B. Let $A = \bigoplus A_i$. There exists a unique homomorphism*

$$f : A \to B ,$$

such that $f \circ \lambda_j = f_j$ for all j.

Proof. We can define a map $f : A \to B$ by the rule

$$f((x_i)_{i \in I}) = \sum_{i \in I} f_i(x_i).$$

The sum on the right is actually finite since all but a finite number of terms are 0. It is immediately verified that our map f is a homomorphism. Furthermore, we clearly have $f \circ \lambda_j(x) = f_j(x)$ for each j and each $x \in A_j$. Thus f has the desired commutativity property. It is also clear that the map f is uniquely determined, as was to be shown.

The property expressed in Proposition 7.1 is called the **universal property** of the direct sum. Cf. §11.

Example. Let A be an abelian group, and let $\{A_i\}_{i \in I}$ be a family of subgroups. Then we get a homomorphism

$$\bigoplus_{i \in I} A_i \to A \quad \text{such that} \quad (x_i) \mapsto \sum x_i.$$

Theorem 8.1 will provide an important specific application.

Let A be an abelian group and B, C subgroups. If $B + C = A$ and $B \cap C = \{0\}$ then the map

$$B \times C \to A$$

given by $(x, y) \mapsto x + y$ is an isomorphism (as we already noted in the non-commutative case). Instead of writing $A = B \times C$ we shall write

$$A = B \oplus C$$

and say that A is the **direct sum** of B and C. We use a similar notation for the direct sum of a finite number of subgroups $B_1, \ldots, B_n$ such that

$$B_1 + \cdots + B_n = A$$

and

$$B_{i+1} \cap (B_1 + \cdots + B_i) = 0.$$

In that case we write

$$A = B_1 \oplus \cdots \oplus B_n.$$

Let A be an abelian group. Let $\{e_i\}$ $(i \in I)$ be a family of elements of A. We say that this family is a **basis** for A if the family is not empty, and if every element of A has a unique expression as a linear combination

$$x = \sum x_i e_i$$

with $x_i \in \mathbf{Z}$ and almost all $x_i = 0$. Thus the sum is actually a finite sum. An abelian group is said to be **free** if it has a basis. If that is the case, it is immediate that if we let $Z_i = \mathbf{Z}$ for all i, then A is isomorphic to the direct sum

$$A \approx \bigoplus_{i \in I} Z_i.$$

Next let S be a set. We shall define the free abelian group generated by S as follows. Let $\mathbf{Z}\langle S\rangle$ be the set of all maps $\varphi : S \to \mathbf{Z}$ such that $\varphi(x) = 0$ for almost all $x \in S$. Then $\mathbf{Z}\langle S\rangle$ is an abelian group (addition being the usual addition of maps). If k is an integer and x is an element of S, we denote by $k \cdot x$ the map φ such that $\varphi(x) = k$ and $\varphi(y) = 0$ if $y \neq x$. Then it is obvious that every element φ of $\mathbf{Z}\langle S\rangle$ can be written in the form

$$\varphi = k_1 \cdot x_1 + \cdots + k_n \cdot x_n$$

for some integers k_i and elements $x_i \in S$ $(i = 1, \ldots, n)$, all the x_i being distinct. Furthermore, *φ admits a unique such expression*, because if we have

$$\varphi = \sum_{x \in S} k_x \cdot x = \sum_{x \in S} k'_x \cdot x$$

then

$$0 = \sum_{x \in S} (k_x - k'_x) \cdot x,$$

whence $k'_x = k_x$ for all $x \in S$.

We map S into $\mathbf{Z}\langle S\rangle$ by the map $f_S = f$ such that $f(x) = 1 \cdot x$. It is then clear that f is injective, and that $f(S)$ generates $\mathbf{Z}\langle S\rangle$. If $g : S \to B$ is a mapping of S into some abelian group B, then we can define a map

$$g_* : \mathbf{Z}\langle S\rangle \to B$$

such that

$$g_*\left(\sum_{x \in S} k_x \cdot x\right) = \sum_{x \in S} k_x g(x).$$

This map is a homomorphism (trivial) and we have $g_* \circ f = g$ (also trivial). It is the only homomorphism which has this property, for any such homomorphism g_* must be such that $g_*(1 \cdot x) = g(x)$.

It is customary to identify S in $\mathbf{Z}\langle S \rangle$, and we sometimes omit the dot when we write $k_x x$ or a sum $\sum k_x x$.

If $\lambda : S \to S'$ is a mapping of sets, there is a unique homomorphism $\bar{\lambda}$ making the following diagram commutative:

$$\begin{array}{ccc} S & \xrightarrow{f_S} & \mathbf{Z}\langle S \rangle \\ \lambda \downarrow & & \downarrow \bar{\lambda} \\ S' & \xrightarrow[f_{S'}]{} & \mathbf{Z}\langle S' \rangle \end{array}$$

In fact, $\bar{\lambda}$ is none other than $(f_{S'} \circ \lambda)_*$, with the notation of the preceding paragraph. The proof of this statement is left as a trivial exercise.

We shall denote $\mathbf{Z}\langle S \rangle$ also by $F_{ab}(S)$, and call $F_{ab}(S)$ the **free abelian group generated by** S. We call elements of S its **free generators**.

As an exercise, show that every abelian group A is a factor group of a free abelian group F. If A is finitely generated, show that one can select F to be finitely generated also.

If the set S above consists of n elements, then we say that the free abelian group $F_{ab}(S)$ is the free abelian group on n generators. If S is the set of n letters $x_1, \dots, x_n$, we say that $F_{ab}(S)$ is the free abelian group with free generators $x_1, \dots, x_n$.

An abelian group is **free** if and only if it is isomorphic to a free abelian group $F_{ab}(S)$ for some set S. Let A be an abelian group, and let S be a basis for A. Then it is clear that A is isomorphic to the free abelian group $F_{ab}(S)$.

As a matter of notation, if A is an abelian group and T a subset of elements of A, we denote by $\langle T \rangle$ the subgroup generated by the elements of T, i.e., the smallest subgroup of A containing T.

Example. The Grothendieck group. Let M be a commutative monoid, written additively. There exists a commutative group $K(M)$ and a monoid-homomorphism

$$\gamma : M \to K(M)$$

having the following universal property. If $f : M \to A$ is a homomorphism into an abelian group A, then there exists a unique homomorphism $f_* : K(M) \to A$ making the following diagram commutative:

$$\begin{array}{ccc} M & \xrightarrow{\gamma} & K(M) \\ & f \searrow \quad \swarrow f_* & \\ & A & \end{array}$$

Proof. Let $F_{ab}(M)$ be the free abelian group generated by M. We denote the generator of $F_{ab}(M)$ corresponding to an element $x \in M$ by $[x]$. Let B be the subgroup generated by all elements of type

$$[x + y] - [x] - [y]$$

where $x, y \in M$. We let $K(M) = F_{ab}(M)/B$, and let

$$\gamma : M \to K(M)$$

be the map obtained by composing the injection of M into $F_{ab}(M)$ given by $x \mapsto [x]$, and the canonical map

$$F_{ab}(M) \to F_{ab}(M)/B.$$

It is then clear that γ is a homomorphism, and satisfies the desired universal property.

The universal group $K(M)$ is called the **Grothendieck group**.

We shall say that the **cancellation law** holds in M if, whenever $x, y, z \in M$, and $x + z = y + z$, we have $x = y$.

We then have an important criterion when the universal map γ above is injective:

If the cancellation law holds in M, then the canonical map γ of M into its Grothendieck group is injective.

Proof. This is essentially the same proof as when one constructs the negative integers from the natural numbers. We consider pairs (x, y) with $x, y \in M$ and say that (x, y) is equivalent to (x', y') if $y + x' = x + y'$. We define addition of pairs componentwise. Then the equivalence classes of pairs form a group, whose 0 element is the class of $(0, 0)$ [or the class of (x, x) for any $x \in M$]. The negative of an element (x, y) is (y, x). We have a homomorphism

$$x \mapsto \text{class of } (0, x)$$

which is injective, as one sees immediately by applying the cancellation law. Thus we have constructed a homomorphism of M into a group, which is injective. It follows that the universal homomorphism must also be injective.

Examples. See the example of projective modules in Chapter III, §4. For a relatively fancy context, see: K. Kato, Logarithmic structures of Fontaine-Illusie, *Algebraic Geometry, Analysis and Number Theory, Proc. JAMI Conference*, J. Igusa (Ed.), Johns Hopkins Press (1989) pp. 195–224.

Given an abelian group A and a subgroup B, it is sometimes desirable to find a subgroup C such that $A = B \oplus C$. The next lemma gives us a condition under which this is true.

Lemma 7.2. *Let $A \xrightarrow{f} A'$ be a surjective homomorphism of abelian groups, and assume that A' is free. Let B be the kernel of f. Then there exists a subgroup C of A such that the restriction of f to C induces an isomorphism of C with A', and such that $A = B \oplus C$.*

Proof. Let $\{x'_i\}_{i \in I}$ be a basis of A', and for each $i \in I$, let x_i be an element of A such that $f(x_i) = x'_i$. Let C be the subgroup of A generated by all elements x_i, $i \in I$. If we have a relation

$$\sum_{i \in I} n_i x_i = 0$$

with integers n_i, almost all of which are equal to 0, then applying f yields

$$0 = \sum_{i \in I} n_i f(x_i) = \sum_{i \in I} n_i x_i',$$

whence all $n_i = 0$. Hence our family $\{x_i\}_{i \in I}$ is a basis of C. Similarly, one sees that if $z \in C$ and $f(z) = 0$ then $z = 0$. Hence $B \cap C = 0$. Let $x \in A$. Since $f(x) \in A'$ there exist integers n_i, $i \in I$, such that

$$f(x) = \sum_{i \in I} n_i x_i'.$$

Applying f to $x - \sum_{i \in I} n_i x_i$, we find that this element lies in the kernel of f, say

$$x - \sum_{i \in I} n_i x_i = b \in B.$$

From this we see that $x \in B + C$, and hence finally that $A = B \oplus C$ is a direct sum, as contended.

Theorem 7.3. *Let A be a free abelian group, and let B be a subgroup. Then B is also a free abelian group, and the cardinality of a basis of B is $\leqq$ the cardinality of a basis for A. Any two bases of B have the same cardinality.*

Proof. We shall give the proof only when A is finitely generated, say by a basis $\{x_1, \ldots, x_n\}$ $(n \geqq 1)$, and give the proof by induction on n. We have an expression of A as direct sum:

$$A = \mathbf{Z}x_1 \oplus \cdots \oplus \mathbf{Z}x_n.$$

Let $f: A \to \mathbf{Z}x_1$ be the projection, i.e. the homomorphism such that

$$f(m_1 x_1 + \cdots + m_n x_n) = m_1 x_1$$

whenever $m_i \in \mathbf{Z}$. Let B_1 be the kernel of $f|B$. Then B_1 is contained in the free subgroup $\langle x_2, \ldots, x_n \rangle$. By induction, B_1 is free and has a basis with $\leqq n - 1$ elements. By the lemma, there exists a subgroup C_1 isomorphic to a subgroup of $\mathbf{Z}x_1$ (namely the image of $f|B$) such that

$$B = B_1 \oplus C_1.$$

Since $f(B)$ is either 0 or infinite cyclic, i.e. free on one generator, this proves that B is free.

(When A is not finitely generated, one can use a similar transfinite argument. See Appendix 2, §2, the example after Zorn's Lemma.)

We also observe that our proof shows that there exists at least one basis of B whose cardinality is $\leqq n$. We shall therefore be finished when we prove the last statement, that any two bases of B have the same cardinality. Let S be one basis, with a finite number of elements m. Let T be another basis, and suppose that T has at least r elements. It will suffice to prove that $r \leqq m$ (one

can then use symmetry). Let p be a prime number. Then B/pB is a direct sum of cyclic groups of order p, with m terms in the sum. Hence its order is p^m. Using the basis T instead of S, we conclude that B/pB contains an r-fold product of cyclic groups of order p, whence $p^r \leqq p^m$, and $r \leqq m$, as was to be shown. (Note that we did not assume a priori that T was finite.)

The number of elements in a basis of a free abelian group A will be called the **rank** of A.

§8. FINITELY GENERATED ABELIAN GROUPS

The groups referred to in the title of this section occur so frequently that it is worth while to state a theorem which describes their structure completely. Throughout this section we write our abelian groups additively.

Let A be an abelian group. An element $a \in A$ is said to be a **torsion** element if it has finite period. The subset of all torsion elements of A is a subgroup of A called the **torsion subgroup** of A. (If a has period m and b has period n then, writing the group law additively, we see that $a \pm b$ has a period dividing mn.)

The torsion subgroup of A is denoted by A_{tor}, or simply A_t. An abelian group is called a **torsion group** if $A = A_{\text{tor}}$, that is all elements of A are of finite order.

A finitely generated torsion abelian group is obviously finite. We shall begin by studying torsion abelian groups. If A is an abelian group and p a prime number, we denote by $A(p)$ the subgroup of all elements $x \in A$ whose period is a power of p. Then $A(p)$ is a torsion group, and is a p-group if it is finite.

Theorem 8.1 *Let A be a torsion abelian group. Then A is the direct sum of its subgroups $A(p)$ for all primes p such that $A(p) \neq 0$.*

Proof. There is a homomorphism

$$\bigoplus_p A(p) \to A$$

which to each element (x_p) in the direct sum associates the element $\sum x_p$ in A. We prove that this homomorphism is both surjective and injective. Suppose x is in the kernel, so $\sum x_p = 0$. Let q be a prime. Then

$$x_q = \sum_{p \neq q} (-x_p).$$

Let m be the least common multiple of the periods of elements x_p on the right-hand side, with $x_q \neq 0$ and $p \neq q$. Then $mx_q = 0$. But also $q^r x_q = 0$ for some positive integer r. If d is the greatest common divisor of m, q^r then $dx_q = 0$, but $d = 1$, so $x_q = 0$. Hence the kernel is trivial, and the homomorphism is injective.

As for the surjectivity, for each positive integer m, denote by A_m the kernel of multiplication by m, i.e. the subgroup of $x \in A$ such that $mx = 0$. We prove:

If $m = rs$ with r, s positive relative prime integers, then $A_m = A_r + A_s$.

Indeed, there exist integers u, v such that $ur + vs = 1$. Then $x = urx + vsx$, and $urx \in A_s$ while $vsx \in A_r$, and our assertion is proved. Repeating this process inductively, we conclude:

$$\textit{If } m = \prod_{p|m} p^{e(p)} \textit{ then } A_m = \sum_{p|m} A_{p^{e(p)}}.$$

Hence the map $\bigoplus A(p) \to A$ is surjective, and the theorem is proved.

Example. Let $A = \mathbf{Q}/\mathbf{Z}$. Then $\mathbf{Q}/\mathbf{Z}$ is a torsion abelian group, isomorphic to the direct sum of its subgroups $(\mathbf{Q}/\mathbf{Z})(p)$. Each $(\mathbf{Q}/\mathbf{Z})(p)$ consists of those elements which can be represented by a rational number a/p^k with $a \in \mathbf{Z}$ and k some positive integer, i.e. a rational number having only a p-power in the denominator. See also Chapter IV, Theorem 5.1.

In what follows we shall deal with finite abelian groups, so only a finite number of primes (dividing the order of the group) will come into play. In this case, the direct sum is "the same as" the direct product.

Our next task is to describe the structure of finite abelian p-groups. Let $r_1, \dots, r_s$ be integers $\geqq 1$. A finite p-group A is said to be of **type** $(p^{r_1}, \dots, p^{r_s})$ if A is isomorphic to the product of cyclic groups of orders p^{r_i} $(i = 1, \dots, s)$. We shall need the following remark.

Remark. Let A be a finite abelian p-group. Let b be an element of A, $b \neq 0$. Let k be an integer $\geqq 0$ such that $p^k b \neq 0$, and let p^m be the period of $p^k b$. Then b has period p^{k+m}. [*Proof*: We certainly have $p^{k+m}b = 0$, and if $p^n b = 0$ then first $n \geqq k$, and second $n \geqq k + m$, otherwise the period of $p^k b$ would be smaller than p^m.]

Theorem 8.2. *Every finite abelian p-group is isomorphic to a product of cyclic p-groups. If it is of type $(p^{r_1}, \dots, p^{r_s})$ with*

$$r_1 \geqq r_2 \geqq \cdots \geqq r_s \geqq 1,$$

then the sequence of integers $(r_1, \dots, r_s)$ is uniquely determined.

Proof. We shall prove the existence of the desired product by induction. Let $a_1 \in A$ be an element of maximal period. We may assume without loss of generality that A is not cyclic. Let A_1 be the cyclic subgroup generated by a_1, say of period p^{r_1}. We need a lemma.

Lemma 8.3. *Let $\bar{b}$ be an element of A/A_1, of period p^r. Then there exists a representative a of $\bar{b}$ in A which also has period p^r.*

Proof. Let b be any representative of $\bar{b}$ in A. Then $p^r b$ lies in A_1, say $p^r b = na_1$ with some integer $n \geqq 0$. We note that the period of $\bar{b}$ is $\leqq$ the period of b. If $n = 0$ we are done. Otherwise write $n = p^k \mu$ where μ is prime to p. Then μa_1 is also a generator of A_1, and hence has period p^{r_1}. We may assume $k \leqq r_1$. Then $p^k \mu a_1$ has period $p^{r_1 - k}$. By our previous remarks, the element b has period

$$p^{r + r_1 - k}$$

whence by hypothesis, $r + r_1 - k \leqq r_1$ and $r \leqq k$. This proves that there exists an element $c \in A_1$ such that $p^r b = p^r c$. Let $a = b - c$. Then a is a representative for $\bar{b}$ in A and $p^r a = 0$. Since period $(a) \leqq p^r$ we conclude that a has period equal to p^r.

We return to the main proof. By induction, the factor group A/A_1 has a product expression

$$A/A_1 = \bar{A}_2 \times \cdots \times \bar{A}_s$$

into cyclic subgroups of orders $p^{r_2}, \dots, p^{r_s}$ respectively, and we may assume $r_2 \geqq \cdots \geqq r_s$. Let $\bar{a}_i$ be a generator for $\bar{A}_i$ $(i = 2, \dots, s)$ and let a_i be a representative in A of the same period as $\bar{a}_i$. Let A_i be the cyclic subgroup generated by a_i. We contend that A is the direct sum of $A_1, \dots, A_s$.

Given $x \in A$, let $\bar{x}$ denote its residue class in A/A_1. There exist integers $m_i \geqq 0$ $(i = 2, \dots, s)$ such that

$$\bar{x} = m_2 \bar{a}_2 + \cdots + m_s \bar{a}_s.$$

Hence $x - m_2 a_2 - \cdots - m_s a_s$ lies in A_1, and there exists an integer $m_1 \geqq 0$ such that

$$x = m_1 a_1 + m_2 a_2 + \cdots + m_s a_s.$$

Hence $A_1 + \cdots + A_s = A$.

Conversely, suppose that $m_1, \dots, m_s$ are integers $\geqq 0$ such that

$$0 = m_1 a_1 + \cdots + m_s a_s.$$

Since a_i has period p^{r_i} $(i = 1, \dots, s)$, we may suppose that $m_i < p^{r_i}$. Putting a bar on this equation yields

$$0 = m_2 \bar{a}_2 + \cdots + m_s \bar{a}_s.$$

Since A/A_1 is a direct product of $\bar{A}_2, \dots, \bar{A}_s$ we conclude that each $m_i = 0$ for $i = 2, \dots, s$. But then $m_1 = 0$ also, and hence all $m_i = 0$ $(i = 1, \dots, s)$. From this it follows at once that

$$(A_1 + \cdots + A_i) \cap A_{i+1} = 0$$

for each $i \geqq 1$, and hence that A is the direct product of $A_1, \dots, A_s$, as desired.

We prove uniqueness, by induction. Suppose that A is written in two ways as a direct sum of cyclic groups, say of type

$$(p^{r_1}, \dots, p^{r_s}) \quad \text{and} \quad (p^{m_1}, \dots, p^{m_k})$$

with $r_1 \geqq \cdots \geqq r_s \geqq 1$ and $m_1 \geqq \cdots \geqq m_k \geqq 1$. Then pA is also a p-group, of order strictly less than the order of A, and is of type

$$(p^{r_1-1}, \ldots, p^{r_s-1}) \quad \text{and} \quad (p^{m_1-1}, \ldots, p^{m_k-1}),$$

it being understood that if some exponent r_i or m_j is equal to 1, then the factor corresponding to

$$p^{r_i-1} \quad \text{or} \quad p^{m_j-1}$$

in pA is simply the trivial group 0. By induction, the subsequence of

$$(r_1 - 1, \ldots, r_s - 1)$$

consisting of those integers $\geqq 1$ is uniquely determined, and is the same as the corresponding subsequence of

$$(m_1 - 1, \ldots, m_k - 1).$$

In other words, we have $r_i - 1 = m_i - 1$ for all those integers i such that $r_i - 1$ or $m_i - 1 \geqq 1$. Hence $r_i = m_i$ for all these integers i, and the two sequences

$$(p^{r_1}, \ldots, p^{r_s}) \quad \text{and} \quad (p^{m_1}, \ldots, p^{m_k})$$

can differ only in their last components which can be equal to p. These correspond to factors of type $(p, \ldots, p)$ occurring say ν times in the first sequences and μ times in the second sequence. Thus for some integer n, A is of type

$$(p^{r_1}, \ldots, p^{r_n}, \underbrace{p, \ldots, p}_{\nu \text{ times}}) \quad \text{and} \quad (p^{r_1}, \ldots, p^{r_n}, \underbrace{p, \ldots, p}_{\mu \text{ times}}).$$

Thus the order of A is equal to

$$p^{r_1 + \cdots + r_n} p^{\nu} = p^{r_1 + \cdots + r_n} p^{\mu},$$

whence $\nu = \mu$, and our theorem is proved.

A group G is said to be **torsion free**, or without torsion, if whenever an element x of G has finite period, then x is the unit element.

Theorem 8.4. *Let A be a finitely generated torsion-free abelian group. Then A is free.*

Proof. Assume $A \neq 0$. Let S be a finite set of generators, and let $x_1, \ldots, x_n$ be a maximal subset of S having the property that whenever $\nu_1, \ldots, \nu_n$ are integers such that

$$\nu_1 x_1 + \cdots + \nu_n x_n = 0,$$

then $\nu_j = 0$ for all j. (Note that $n \geqq 1$ since $A \neq 0$). Let B be the subgroup generated by $x_1, \ldots, x_n$. Then B is free. Given $y \in A$ there exist integers $m_1, \ldots, m_n, m$ not all zero such that

$$my + m_1 x_1 + \cdots + m_n x_n = 0,$$

by the assumption of maximality on $x_1, \dots, x_n$. Furthermore, $m \neq 0$; otherwise all $m_j = 0$. Hence my lies in B. This is true for every one of a finite set of generators y of A, whence there exists an integer $m \neq 0$ such that $mA \subset B$. The map

$$x \mapsto mx$$

of A into itself is a homomorphism, having trivial kernel since A is torsion free. Hence it is an isomorphism of A onto a subgroup of B. By Theorem 7.3 of the preceding section, we conclude that mA is free, whence A is free.

Theorem 8.5. *Let A be a finitely generated abelian group, and let A_{tor} be the subgroup consisting of all elements of A having finite period. Then A_{tor} is finite, and A/A_{tor} is free. There exists a free subgroup B of A such that A is the direct sum of A_{tor} and B.*

Proof. We recall that a finitely generated torsion abelian group is obviously finite. Let A be finitely generated by n elements, and let F be the free abelian group on n generators. By the universal property, there exists a surjective homomorphism

$$F \xrightarrow{\varphi} A$$

of F onto A. The subgroup $\varphi^{-1}(A_{\text{tor}})$ of F is finitely generated by Theorem 7.3. Hence A_{tor} itself is finitely generated, hence finite.

Next, we prove that A/A_{tor} has no torsion. Let $\bar{x}$ be an element of A/A_{tor} such that $m\bar{x} = 0$ for some integer $m \neq 0$. Then for any representative of x of $\bar{x}$ in A, we have $mx \in A_{\text{tor}}$, whence $qmx = 0$ for some integer $q \neq 0$. Then $x \in A_{\text{tor}}$, so $\bar{x} = 0$, and A/A_{tor} is torsion free. By Theorem 8.4, A/A_{tor} is free. We now use the lemma of Theorem 7.3 to conclude the proof.

The rank of A/A_{tor} is also called the **rank** of A.

For other contexts concerning Theorem 8.5, see the structure theorem for modules over principal rings in Chapter III, §7, and Exercises 5, 6, and 7 of Chapter III.

§9. THE DUAL GROUP

Let A be an abelian group of exponent $m \geqq 1$. This means that for each element $x \in A$ we have $mx = 0$. Let Z_m be a cyclic group of order m. We denote by $A^\wedge$, or $\text{Hom}(A, Z_m)$ the group of homomorphisms of A into Z_m, and call it the **dual** of A.

Let $f : A \to B$ be a homomorphism of abelian groups, and assume both have exponent m. Then f induces a homomorphism

$$f^\wedge : B^\wedge \to A^\wedge.$$

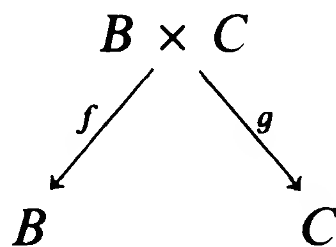
Namely, for each $\psi \in B^\wedge$ we define $f^\wedge(\psi) = \psi \circ f$. It is trivially verified that $f^\wedge$ is a homomorphism. The properties

$$\mathrm{id}^\wedge = \mathrm{id} \quad \text{and} \quad (f \circ g)^\wedge = g^\wedge \circ f^\wedge$$

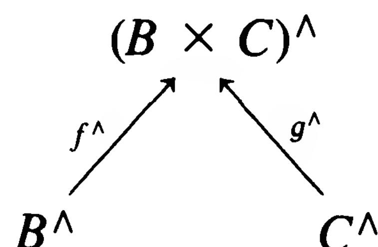
are trivially verified.

Theorem 9.1. *If A is a finite abelian group, expressed as a product $A = B \times C$, then $A^\wedge$ is isomorphic to $B^\wedge \times C^\wedge$ (under the mapping described below). A finite abelian group is isomorphic to its own dual.*

Proof. Consider the two projections



of $B \times C$ on its two components. We get homomorphisms



and we contend that these homomorphisms induce an isomorphism of $B^\wedge \times C^\wedge$ onto $(B \times C)^\wedge$.

In fact, let ψ_1, ψ_2 be in $\mathrm{Hom}(B, Z_m)$ and $\mathrm{Hom}(C, Z_m)$ respectively. Then $(\psi_1, \psi_2) \in B^\wedge \times C^\wedge$, and we have a corresponding element of $(B \times C)^\wedge$ by defining

$$(\psi_1, \psi_2)(x, y) = \psi_1(x) + \psi_2(y),$$

for $(x, y) \in B \times C$. In this way we get a homomorphism

$$B^\wedge \times C^\wedge \to (B \times C)^\wedge.$$

Conversely, let $\psi \in (B \times C)^\wedge$. Then

$$\psi(x, y) = \psi(x, 0) + \psi(0, y).$$

The function ψ_1 on B such that $\psi_1(x) = \psi(x, 0)$ is in $B^\wedge$, and similarly the function ψ_2 on C such that $\psi_2(y) = \psi(0, y)$ is in $C^\wedge$. Thus we get a homomorphism

$$(B \times C)^\wedge \to B^\wedge \times C^\wedge,$$

which is obviously inverse to the one we defined previously. Hence we obtain an isomorphism, thereby proving the first assertion in our theorem.

We can write any finite abelian group as a product of cyclic groups. Thus to prove the second assertion, it will suffice to deal with a cyclic group.

Let A be cyclic, generated by one element x of period n. Then $n \mid m$, and Z_m has precisely one subgroup of order n, Z_n, which is cyclic (Proposition 4.3(**iv**)).

If $\psi : A \to Z_m$ is a homomorphism, and x is a generator for A, then the period of x is an exponent for $\psi(x)$, so that $\psi(x)$, and hence $\psi(A)$, is contained in Z_n. Let y be a generator for Z_n. We have an isomorphism

$$\psi_1 : A \to Z_n$$

such that $\psi_1(x) = y$. For each integer k with $0 \leqq k < n$ we have the homomorphism $k\psi_1$ such that

$$(k\psi_1)(x) = k \cdot \psi_1(x) = \psi_1(kx).$$

In this way we get a cyclic subgroup of $A^\wedge$ consisting of the n elements $k\psi_1$ $(0 \leqq k < n)$. Conversely, any element ψ of $A^\wedge$ is uniquely determined by its effect on the generator x, and must map x on one of the n elements kx $(0 \leqq k < n)$ of Z_n. Hence ψ is equal to one of the maps $k\psi_1$. These maps constitute the full group $A^\wedge$, which is therefore cyclic of order n, generated by ψ_1. This proves our theorem.

In considering the dual group, we take various cyclic groups Z_m. There are many applications where such groups occur, for instance the group of m-th roots of unity in the complex numbers, the subgroup of order m of $\mathbf{Q}/\mathbf{Z}$, etc.

Let A, A' be two abelian groups. A **bilinear** map of $A \times A'$ into an abelian group C is a map

$$A \times A' \to C$$

denoted by

$$(x, x') \mapsto \langle x, x' \rangle$$

having the following property. For each $x \in A$ the function $x' \mapsto \langle x, x' \rangle$ is a homomorphism, and similarly for each $x' \in A'$ the function $x \mapsto \langle x, x' \rangle$ is a homomorphism.

As a special case of a bilinear map, we have the one given by

$$A \times \mathrm{Hom}(A, C) \to C$$

which to each pair (x, f) with $x \in A$ and $f \in \mathrm{Hom}(A, C)$ associates the element $f(x)$ in C.

A bilinear map is also called a **pairing**.

An element $x \in A$ is said to be **orthogonal** (or **perpendicular**) to a subset S' of A' if $\langle x, x' \rangle = 0$ for all $x' \in S'$. It is clear that the set of $x \in A$ orthogonal to S' is a subgroup of A. We make similar definitions for elements of A', orthogonal to subsets of A.

The **kernel** of our bilinear map on the left is the subgroup of A which is orthogonal to all of A'. We define its kernel on the right similarly.

Given a bilinear map $A \times A' \to C$, let B, B' be the respective kernels of our bilinear map on the left and right. An element x' of A' gives rise to an element of $\mathrm{Hom}(A, C)$ given by $x \mapsto \langle x, x' \rangle$, which we shall denote by $\psi_{x'}$. Since $\psi_{x'}$ vanishes on B we see that $\psi_{x'}$ is in fact a homomorphism of A/B into C.

Furthermore, $\psi_{x'} = \psi_{y'}$ if x', y' are elements of A' such that

$$x' \equiv y' \pmod{B'}.$$

Hence ψ is in fact a homomorphism

$$0 \to A'/B' \to \mathrm{Hom}(A/B, C),$$

which is injective since we defined B' to be the group orthogonal to A. Similarly, we get an injective homomorphism

$$0 \to A/B \to \mathrm{Hom}(A'/B', C).$$

Assume that C is cyclic of order m. Then for any $x' \in A'$ we have

$$m\psi_{x'} = \psi_{mx'} = 0,$$

whence A'/B' has exponent m. Similarly, A/B has exponent m.

Theorem 9.2. *Let $A \times A' \to C$ be a bilinear map of two abelian groups into a cyclic group C of order m. Let B, B' be its respective kernels on the left and right. Assume that A'/B' is finite. Then A/B is finite, and A'/B' is isomorphic to the dual group of A/B (under our map ψ).*

Proof. The injection of A/B into $\mathrm{Hom}(A'/B', C)$ shows that A/B is finite. Furthermore, we get the inequalities

$$\operatorname{ord} A/B \leqq \operatorname{ord}(A'/B')^\wedge = \operatorname{ord} A'/B'$$

and

$$\operatorname{ord} A'/B' \leqq \operatorname{ord}(A/B)^\wedge = \operatorname{ord} A/B.$$

From this it follows that our map ψ is bijective, hence an isomorphism.

Corollary 9.3. *Let A be a finite abelian group, B a subgroup, $A^\wedge$ the dual group, and $B^\perp$ the set of $\varphi \in A^\wedge$ such that $\varphi(B) = 0$. Then we have a natural isomorphism of $A^\wedge/B^\perp$ with $B^\wedge$.*

Proof. This is a special case of Theorem 9.2.

§10. INVERSE LIMIT AND COMPLETION

Consider a sequence of groups $\{G_n\}$ $(n = 0, 1, 2, \ldots)$, and suppose given for all $n \geqq 1$ homomorphisms

$$f_n \colon G_n \to G_{n-1}.$$

Suppose first that these homomorphisms are surjective. We form infinite sequences

$$x = (x_0, x_1, x_2, \ldots) \quad \text{such that } x_{n-1} = f_n(x_n).$$

By the assumption of surjectivity, given $x_n \in G_n$ we can always lift x_n to G_{n+1} via f_{n+1}, so such infinite sequences exist, projecting to any given x_0. We can define multiplication of such sequences componentwise, and it is then immediately verified that the set of sequences is a group, called the **inverse limit** of the family $\{(G_n, f_n)\}$. We denote the inverse limit by $\varprojlim (G_n, f_n)$, or simply $\lim G_n$ if the reference to f_n is clear.

Example. Let A be an additive abelian group. Let p be a prime number. Let $p_A: A \to A$ denote multiplication by p. We say that A is **p-divisible** if p_A is surjective. We may then form the inverse limit by taking $A_n = A$ for all n, and $f_n = p_A$ for all n. The inverse limit is denoted by $V_p(A)$. We let $T_p(A)$ be the subset of $V_p(A)$ consisting of those infinite sequences as above such that $x_0 = 0$. Let $A[p^n]$ be the kernel of p_A^n. Then

$$T_p(A) = \lim A[p^{n+1}].$$

The group $T_p(A)$ is called the **Tate group** associated with the p-divisible group A. It arose in fairly sophisticated contexts of algebraic geometry due to Deuring and Weil, in the theory of elliptic curves and abelian varieties developed in the 1940s, which are far afield from this book. Interested readers can consult books on those subjects.

The most common p-divisible groups are obtained as follows. First, let A be the subgroup of $\mathbf{Q}/\mathbf{Z}$ consisting of those rational numbers (mod $\mathbf{Z}$) which can be expressed in the form a/p^k with some positive integer k, and $a \in \mathbf{Z}$. Then A is p-divisible.

Second, let $\boldsymbol{\mu}[p^n]$ be the group of p^n-th roots of unity in the complex numbers. Let $\boldsymbol{\mu}[p^\infty]$ be the union of all $\boldsymbol{\mu}[p^n]$ for all n. Then $\boldsymbol{\mu}[p^\infty]$ is p-divisible, and isomorphic to the group A of the preceding paragraph. Thus

$$T_p(\boldsymbol{\mu}) = \lim \boldsymbol{\mu}[p^n].$$

These groups are quite important in number theory and algebraic geometry. We shall make further comments about them in Chapter III, §10, in a broader context.

Example. Suppose given a group G. Let $\{H_n\}$ be a sequence of normal subgroups such that $H_n \supset H_{n+1}$ for all n. Let

$$f_n: G/H_n \to G/H_{n-1}$$

be the canonical homomorphisms. Then we may form the inverse limit $\varprojlim G/H_n$. Observe that G has a natural homomorphism

$$g: G \to \varprojlim G/H_n,$$

which sends an element x to the sequence $(\ldots, x_n, \ldots)$, where $x_n =$ image of x in G/H_n.

Example. Let $G_n = \mathbf{Z}/p^{n+1}\mathbf{Z}$ for each $n \geqq 0$. Let

$$f_n: \mathbf{Z}/p^{n+1}\mathbf{Z} \to \mathbf{Z}/p^n\mathbf{Z}$$

be the canonical homomorphism. Then f_n is surjective, and the limit is called

the group of p-**adic integers**, denoted by $\mathbf{Z}_p$. We return to this in Chapter III, §10, where we shall see that $\mathbf{Z}_p$ is also a ring.

After these examples, we want to consider the more general situation when one deals not with a sequence but with a more general type of family of groups, which may not be commutative. We therefore define inverse limits of groups in general.

Let I be a set of indices. Suppose given a relation of partial ordering in I, namely for some pairs (i, j) we have a relation $i \leqq j$ satisfying the conditions: For all i, j, k in I, we have $i \leqq i$; if $i \leqq j$ and $j \leqq k$ then $i \leqq k$; if $i \leqq j$ and $j \leqq i$ then $i = j$. We say that I is **directed** if given $i, j \in I$, there exists k such that $i \leqq k$ and $j \leqq k$. Assume that I is directed. By an (inversely) **directed family** of groups, we mean a family $\{G_i\}_{i \in I}$ and for each pair $i \leqq j$ a homomorphism

$$f_i^j : G_j \to G_i$$

such that, whenever $k \leqq i \leqq j$ we have

$$f_k^i \circ f_i^j = f_k^j \quad \text{and} \quad f_i^i = \text{id}.$$

Let $G = \prod G_i$ be the product of the family. Let Γ be the subset of G consisting of all elements (x_i) with $x_i \in G_i$ such that for all i and $j \geqq i$ we have

$$f_i^j(x_j) = x_i.$$

Then Γ contains the unit element, and is immediately verified to be a subgroup of G. We call Γ the **inverse limit** of the family, and write

$$\Gamma = \varprojlim G_i.$$

Example. Let G be a group. Let $\mathfrak{F}$ be the family of normal subgroups of finite index. If H, K are normal of finite index, then so is $H \cap K$, so $\mathfrak{F}$ is a directed family. We may then form the inverse limit $\varprojlim G/H$ with $H \in \mathfrak{F}$. There is a variation on this theme. Instead of $\mathfrak{F}$, let p be a prime number, and let $\mathfrak{F}_p$ be the family of normal subgroups of finite index equal to a power of p. Then the inverse limit with respect to subgroups $H \in \mathfrak{F}_p$ can also be taken. (Verify that if H, K are normal of finite p-power index, so is their intersection.)

A group which is an inverse limit of finite groups is called **profinite**.

Example from applications. Such inverse limits arise in Galois theory. Let k be a field and let A be an infinite Galois extension. For example, $k = \mathbf{Q}$ and A is an algebraic closure of $\mathbf{Q}$. Let G be the Galois group; that is, the group of automorphisms of A over k. Then G is the inverse limit of the factor groups G/H, where H ranges over the Galois groups of A over K, with K ranging over all finite extensions of k contained in A. See the Shafarevich conjecture in the chapter on Galois theory, Conjecture 14.2 of Chapter VI.

Similarly, consider a compact Riemann surface X of genus $\geqq 2$. Let $p : X' \to X$ be the universal covering space. Let $\mathbf{C}(X) = F$ and $\mathbf{C}(X') = F'$ be the function fields. Then there is an embedding $\pi_1(X) \hookrightarrow \text{Gal}(F'/F)$. It is shown in complex analysis that $\pi_1(X)$ is a free group with one commutator

relation. The full Galois group of F'/F is the inverse limit with respect to the subgroups of finite index, as in the above general situation.

Completion of a group

Suppose now that we are given a group G, and first, for simplicity, suppose given a sequence of normal subgroups $\{H_r\}$ with $H_r \supset H_{r+1}$ for all n, and such that these subgroups have finite index. A sequence $\{x_n\}$ in G will be called a **Cauchy sequence** if given H_r there exists N such that for all $m, n \geqq N$ we have $x_n x_m^{-1} \in H_r$. We say that $\{x_n\}$ is a **null sequence** if given r there exists N such that for all $n \geqq N$ we have $x_n \in H_r$. As an exercise, prove that the Cauchy sequences form a group under termwise product, and that the null sequences form a normal subgroup. The factor group is called the **completion** of G (with respect to the sequence of normal subgroups).

Observe that there is a natural homomorphism of G into its completion; namely, an element $x \in G$ maps to the sequence $(x, x, x, \ldots)$ modulo null sequences. The kernel of this homomorphism is the intersection $\bigcap H_r$, so if this intersection is the unit element of G, then the map of G into its completion is an embedding.

Theorem 10.1. *The completion and the inverse limit* $\varprojlim G/H_r$ *are isomorphic under natural mappings.*

Proof. We give the maps. Let $x = \{x_n\}$ be a Cauchy sequence. Given r, for all n sufficiently large, by the definition of Cauchy sequence, the class of x_n mod H_r is independent of n. Let this class be $x(r)$. Then the sequence $(x(1), x(2), \ldots)$ defines an element of the inverse limit. Conversely, given an element $(\bar{x}_1, \bar{x}_2, \ldots)$ in the inverse limit, with $\bar{x}_n \in G/H_n$, let x_n be a representative in G. Then the sequence $\{x_n\}$ is Cauchy. We leave to the reader to verify that the Cauchy sequence $\{x_n\}$ is well-defined modulo null sequences, and that the maps we have defined are inverse isomorphisms between the completion and the direct limit.

We used sequences and denumerability to make the analogy with the construction of the real numbers clearer. In general, given the family $\mathfrak{F}$, one considers families $\{x_H\}_{H \in \mathfrak{F}}$ of elements $x_H \in G$. Then the condition for a **Cauchy family** reads: given $H_0 \in \mathfrak{F}$ there exists $H_1 \in \mathfrak{F}$ such that if K, K' are contained in H_1, then $x_K x_{K'}^{-1} \in H_0$. In practice, one can work with sequences, because groups that arise naturally are such that the set of subgroups of finite index is denumerable. This occurs when the group G is countably generated.

More generally, a family $\{H_i\}$ of normal subgroups of finite index is called **cofinal** if given $H \in \mathfrak{F}$ there exists i such that $H_i \subset H$. Suppose that there exists such a family which is denumerable; that is, $i = 1, 2, \ldots$ ranges over the positive integers. Then it is an exercise to show that there is an isomorphism

$$\varprojlim_i G/H_i \approx \varprojlim_{H \in \mathfrak{F}} G/H,$$

or equivalently, that the completion of G with respect to the sequence $\{H_i\}$ is "the same" as the completion with respect to the full family $\mathfrak{F}$. We leave this verification to the reader.

The process of completion is frequent in mathematics. For instance, we shall mention completions of rings in Chapter III, §10; and in Chapter XII we shall deal with completions of fields.

§11. CATEGORIES AND FUNCTORS

Before proceeding further, it will now be convenient to introduce some new terminology. We have met already several kinds of objects: sets, monoids, groups. We shall meet many more, and for each such kind of objects we define special kinds of maps between them (e.g. homomorphisms). Some formal behavior will be common to all of these, namely the existence of identity maps of an object onto itself, and the associativity of maps when such maps occur in succession. We introduce the notion of category to give a general setting for all of these.

A **category** $\mathfrak{A}$ consists of a collection of objects $\text{Ob}(\mathfrak{A})$; and for two objects $A, B \in \text{Ob}(\mathfrak{A})$ a set $\text{Mor}(A, B)$ called the set of **morphisms** of A into B; and for three objects $A, B, C \in \text{Ob}(\mathfrak{A})$ a law of composition (i.e. a map)

$$\text{Mor}(B, C) \times \text{Mor}(A, B) \to \text{Mor}(A, C)$$

satisfying the following axioms:

CAT 1. Two sets $\text{Mor}(A, B)$ and $\text{Mor}(A', B')$ are disjoint unless $A = A'$ and $B = B'$, in which case they are equal.

CAT 2. For each object A of $\mathfrak{A}$ there is a morphism $\text{id}_A \in \text{Mor}(A, A)$ which acts as left and right identity for the elements of $\text{Mor}(A, B)$ and $\text{Mor}(B, A)$ respectively, for all objects $B \in \text{Ob}(\mathfrak{A})$.

CAT 3. The law of composition is associative (when defined), i.e. given $f \in \text{Mor}(A, B)$, $g \in \text{Mor}(B, C)$ and $h \in \text{Mor}(C, D)$ then

$$(h \circ g) \circ f = h \circ (g \circ f),$$

for all objects A, B, C, D of $\mathfrak{A}$.

Here we write the composition of an element g in $\text{Mor}(B, C)$ and an element f in $\text{Mor}(A, B)$ as $g \circ f$, to suggest composition of mappings. In practice, in this book we shall see that most of our morphisms are actually mappings, or closely related to mappings.

The collection of all morphisms in a category $\mathfrak{A}$ will be denoted by $\text{Ar}(\mathfrak{A})$ ("arrows of $\mathfrak{A}$"). We shall sometimes use the symbols "$f \in \text{Ar}(\mathfrak{A})$" to mean

that f is a morphism of $\mathfrak{A}$, i.e. an element of some set $\mathrm{Mor}(A, B)$ for some $A, B \in \mathrm{Ob}(\mathfrak{A})$.

By abuse of language, we sometimes refer to the collection of objects as the category itself, if it is clear what the morphisms are meant to be.

An element $f \in \mathrm{Mor}(A, B)$ is also written $f: A \to B$ or

$$A \xrightarrow{f} B.$$

A morphism f is called an **isomorphism** if there exists a morphism $g: B \to A$ such that $g \circ f$ and $f \circ g$ are the identities in $\mathrm{Mor}(A, A)$ and $\mathrm{Mor}(B, B)$ respectively. If $A = B$, then we also say that the isomorphism is an **automorphism**.

A morphism of an object A into itself is called an **endomorphism**. The set of endomorphisms of A is denoted by $\mathrm{End}(A)$. It follows at once from our axioms that $\mathrm{End}(A)$ is a monoid.

Let A be an object of a category $\mathfrak{A}$. We denote by $\mathrm{Aut}(A)$ the set of automorphisms of A. This set is in fact a group, because all of our definitions are so adjusted so as to see immediately that the group axioms are satisfied (associativity, unit element, and existence of inverse). Thus we now begin to see some feedback between abstract categories and more concrete ones.

Examples. Let $\mathcal{S}$ be the category whose objects are sets, and whose morphisms are maps between sets. We say simply that $\mathcal{S}$ is the category of sets. The three axioms **CAT 1, 2, 3** are trivially satisfied.

Let **Grp** be the category of groups, i.e. the category whose objects are groups and whose morphisms are group-homomorphisms. Here again the three axioms are trivially satisfied. Similarly, we have a category of monoids, denoted by **Mon**.

Later, when we define rings and modules, it will be clear that rings form a category, and so do modules over a ring.

It is important to emphasize here that there are categories for which the set of morphisms is not an abelian group. Some of the most important examples are:

The category $\mathcal{C}^0$, whose objects are open sets in $\mathbf{R}^n$ and whose morphisms are continuous maps.

The category $\mathcal{C}^\infty$ with the same objects, but whose morphisms are the C^∞ maps.

The category **Hol**, whose objects are open sets in $\mathbf{C}^n$, and whose morphisms are holomorphic maps. In each case the axioms of a category are verified, because for instance for **Hol**, the composite of holomorphic maps is holomorphic, and similarly for the other types of maps. Thus a C^0-isomorphism is a continuous map $f: U \to V$ which has a continuous inverse $g: V \to U$. Note that a map may be a C^0-isomorphism but not a C^∞-isomorphism. For instance, $x \mapsto x^3$ is a C^0-automorphism of $\mathbf{R}$, but its inverse is not differentiable.

In mathematics one studies manifolds in any one of the above categories. The determination of the group of automorphisms in each category is one of the basic problems of the area of mathematics concerned with that category. In

complex analysis, one determines early the group of holomorphic automorphisms of the unit disc as the group of all maps

$$z \mapsto e^{i\theta} \frac{c - z}{1 - \bar{c}z}$$

with θ real and $c \in \mathbf{C}$, $|c| < 1$.

Next we consider the notion of operation in categories. First, observe that if G is a group, then the G-sets form a category, whose morphisms are the maps $f : S \to S'$ such that $f(xs) = xf(s)$ for $x \in G$ and $s \in S$.

More generally, we can now define the notion of an operation of a group G on an object in any category. Indeed, let $\mathfrak{A}$ be a category and $A \in \mathrm{Ob}(\mathfrak{A})$. By an **operation** of G on A we shall mean a homomorphism of G into the group $\mathrm{Aut}(A)$. In practice, an object A is a set with elements, and an automorphism in $\mathrm{Aut}(A)$ operates on A as a set, i.e. induces a permutation of A. Thus, if we have a homomorphism

$$\rho : G \to \mathrm{Aut}(A),$$

then for each $x \in G$ we have an automorphism $\rho(x)$ of A which is a permutation of A.

An operation of a group G on an object A is also called a **representation** of G on A, and one then says that G is **represented** as a group of automorphisms of A.

Examples. One meets representations in many contexts. In this book, we shall encounter representations of a group on finite-dimensional vector spaces, with the theory pushed to some depth in Chapter XVIII. We shall also deal with representations of a group on modules over a ring. In topology and differential geometry, one represents groups as acting on various topological spaces, for instance spheres. Thus if X is a differential manifold, or a topological manifold, and G is a group, one considers all possible homomorphims of G into $\mathrm{Aut}(X)$, where Aut refers to whatever category is being dealt with. Thus G may be represented in the group of C^0-automorphims, or C^∞-automorphisms, or analytic automorphisms. Such topological theories are not independent of the algebraic theories, because by functoriality, an action of G on the manifold induces an action on various algebraic functors (homology, K-functor, whatever), so that topological or differential problems are to some extent analyzable by the functorial action on the associated groups, vector spaces, or modules.

Let A, B be objects of a category $\mathfrak{A}$. Let $\mathrm{Iso}(A, B)$ be the set of isomorphisms of A with B. Then the group $\mathrm{Aut}(B)$ operates on $\mathrm{Iso}(A, B)$ by composition; namely, if $u \in \mathrm{Iso}(A, B)$ and $v \in \mathrm{Aut}(B)$, then $(v, u) \mapsto v \circ u$ gives the operation. If u_0 is one element of $\mathrm{Iso}(A, B)$, then the orbit of u_0 is all of $\mathrm{Iso}(A, B)$, so $v \mapsto v \circ u_0$ is a bijection $\mathrm{Aut}(B) \to \mathrm{Iso}(A, B)$. The inverse mapping is given by $u \mapsto u_0\, u_0^{-1}$. This trivial formalism is very basic, and is applied constantly to each one of the classical categories mentioned above. Of course, we also have

a similar bijection on the other side, but the group Aut(A) operates *on the right* of Iso(A, B) by composition. Furthermore, if $u: A \to B$ is an isomorphism, then Aut(A) and Aut(B) are isomorphic under conjugation, namely

$$w \mapsto uwu^{-1} \quad \text{is an isomorphism} \quad \mathrm{Aut}(A) \to \mathrm{Aut}(B).$$

Two such isomorphisms differ by an inner automorphism. One may visualize this system via the following commutative diagram.

$$\begin{array}{ccc} A & \xrightarrow{u} & B \\ {\scriptstyle w}\downarrow & & \downarrow{\scriptstyle uwu^{-1}} \\ A & \xrightarrow[u]{} & B \end{array}$$

Let $\rho : G \to \mathrm{Aut}(A)$ and $\rho' : G \to \mathrm{Aut}(A')$ be representations of a group G on two objects A and A' in the same category. A **morphism** of ρ into ρ' is a morphism $h: A \to A'$ such that the following diagram is commutative for all $x \in G$:

$$\begin{array}{ccc} A & \xrightarrow{h} & A' \\ {\scriptstyle \rho(x)}\downarrow & & \downarrow{\scriptstyle \rho'(x)} \\ A & \xrightarrow[h]{} & A' \end{array}$$

It is then clear that representations of a group G in the objects of a category $\mathfrak{A}$ themselves form a category. An **isomorphism of representations** is then an isomorphism $h : A \to A'$ making the above diagram commutative. An isomorphism of representations is often called an equivalence, but I don't like to tamper with the general system of categorical terminology. Note that if h is an isomorphism of representations, then instead of the above commutative diagram, we let $[h]$ be conjugation by h, and we may use the equivalent diagram

$$\begin{array}{ccc} & \nearrow^{\rho} & \mathrm{Aut}(A) \\ G & & \downarrow{\scriptstyle [h]} \\ & \searrow_{\rho'} & \mathrm{Aut}(A') \end{array}$$

Consider next the case where $\mathfrak{A}$ is the category of abelian groups, which we may denote by **Ab**. Let A be an abelian group and G a group. Given an operation of G on the abelian group A, i.e. a homomorphism

$$\rho : G \to \mathrm{Aut}(A),$$

let us denote by $x \cdot a$ the element $\rho_x(a)$. Then we see that for all $x, y \in G$, a, $b \in A$, we have:

$$x \cdot (y \cdot a) = (xy) \cdot a, \qquad x \cdot (a + b) = x \cdot a + x \cdot b,$$
$$e \cdot a = a, \qquad x \cdot 0 = 0.$$

We observe that when a group G operates on itself by conjugation, then not only does G operate on itself as a set but also operates on itself as an object in the category of groups, i.e. the permutations induced by the operation are actually group-automorphisms.

Similarly, we shall introduce later other categories (rings, modules, fields) and we have given a general definition of what it means for a group to operate on an object in any one of these categories.

Let $\mathfrak{A}$ be a category. We may take as objects of a new category $\mathfrak{C}$ the morphisms of $\mathfrak{A}$. If $f: A \to B$ and $f': A' \to B'$ are two morphisms in $\mathfrak{A}$ (and thus objects of $\mathfrak{C}$), then we define a **morphism** $f \to f'$ (in $\mathfrak{C}$) to be a pair of morphisms (φ, ψ) in $\mathfrak{A}$ making the following diagram commutative:

$$\begin{array}{ccc} A & \xrightarrow{f} & B \\ \varphi \downarrow & & \downarrow \psi \\ A' & \xrightarrow{f'} & B' \end{array}$$

In that way, it is clear that $\mathfrak{C}$ is a category. Strictly speaking, as with maps of sets, we should index (φ, ψ) by f and f' (otherwise **CAT 1** is not necessarily satisfied), but such indexing is omitted in practice.

There are many variations on this example. For instance, we could restrict our attention to morphisms in $\mathfrak{A}$ which have a fixed object of departure, or those which have a fixed object of arrival.

Thus let A be an object of $\mathfrak{A}$, and let $\mathfrak{A}_A$ be the category whose objects are morphisms

$$f: X \to A$$

in $\mathfrak{A}$, having A as object of arrival. A morphism in $\mathfrak{A}_A$ from $f: X \to A$ to $g: Y \to A$ is simply a morphism

$$h: X \to Y$$

in $\mathfrak{A}$ such that the diagram is commutative:

$$\begin{array}{ccc} X & \xrightarrow{h} & Y \\ & f \searrow \quad \swarrow g & \\ & A & \end{array}$$

Universal objects

Let $\mathfrak{C}$ be a category. An object P of $\mathfrak{C}$ is called **universally attracting** if there exists a unique morphism of each object of $\mathfrak{C}$ into P, and is called **universally repelling** if for every object of $\mathfrak{C}$ there exists a unique morphism of P into this object.

When the context makes our meaning clear, we shall call objects P as above **universal**. Since a universal object P admits the identity morphism into itself, it is clear that if P, P' are two universal objects in $\mathcal{C}$, then there exists a unique isomorphism between them.

Examples. Note that the trivial group consisting only of one element is universal (repelling and attracting) in the category of groups. Similarly, in Chapter II on rings, you will see that the integers **Z** are universal in the category of rings (universally repelling).

Next let S be a set. Let $\mathcal{C}$ be the category whose objects are maps $f: S \to A$ of S into abelian groups, and whose morphisms are the obvious ones: If $f: S \to A$ and $f': S \to A'$ are two maps into abelian groups, then a morphism of f into f' is a (group) homomorphism $g: A \to A'$ such that the usual diagram is commutative, namely $g \circ f = f'$. Then the free abelian group generated by S is universal in this category. This is a reformulation of the properties we have proved about this group.

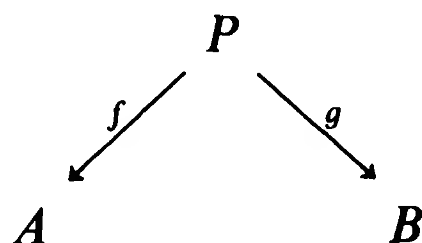
Let M be a commutative monoid and let γ: $M \to K(M)$ be the canonical homomorphism of M into its Grothendieck group. Then γ is universal in the category of homomorphisms of M into abelian groups.

Throughout this book in numerous situtaions, we define universal objects. Aside from products and coproducts which come immediately after these examples, we have direct and inverse limits; the tensor product in Chapter XVI, §1; the alternating product in Chapter XIX, §1; Clifford algebras in Chapter XIX, §4; *ad lib.*

We now turn to the notion of product in an arbitrary category.

Products and coproducts

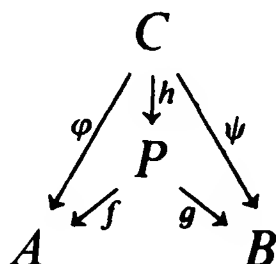
Let $\mathfrak{A}$ be a category and let A, B be objects of $\mathfrak{A}$. By a **product** of A, B in $\mathfrak{A}$ one means a triple (P, f, g) consisting of an object P in $\mathfrak{A}$ and two morphisms



satisfying the following condition: Given two morphisms

$$\varphi: C \to A \quad \text{and} \quad \psi: C \to B$$

in $\mathfrak{A}$, there exists a unique morphism $h: C \to P$ which makes the following diagram commutative:



In other words, $\varphi = f \circ h$ and $\psi = g \circ h$.

More generally, given a family of objects $\{A_i\}_{i \in I}$ in $\mathfrak{A}$, a **product** for this family consists of $(P, \{f_i\}_{i \in I})$, where P is an object in $\mathfrak{A}$ and $\{f_i\}_{i \in I}$ is a family of morphisms

$$f_i : P \to A_i,$$

satisfying the following condition: Given a family of morphisms

$$g_i : C \to A_i,$$

there exists a unique morphism $h : C \to P$ such that $f_i \circ h = g_i$ for all i.

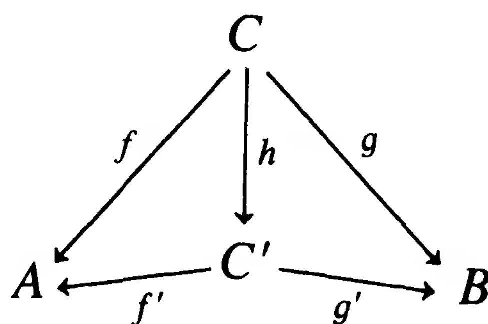
Example. Let $\mathfrak{A}$ be the category of sets, and let $\{A_i\}_{i \in I}$ be a family of sets. Let $A = \prod_{i \in I} A_i$ be their cartesian product, and let $p_i : A \to A_i$ be the projection on the i-th factor. Then $(A, \{p_i\})$ clearly satisfies the requirements of a product in the category of sets.

As a matter of notation, we shall usually write $A \times B$ for the product of two objects in a category, and $\prod_{i \in I} A_i$ for the product of an arbitrary family in a category, following the same notation as in the category of sets.

Example. *Let $\{G_i\}_{i \in I}$ be a family of groups, and let $G = \prod G_i$ be their direct product. Let $p_i : G \to G_i$ be the projection homomorphism. Then these constitute a product of the family in the category of groups.*

Indeed, if $\{g_i : G' \to G_i\}_{i \in I}$ is a family of homomorphisms, there is a unique homomorphism $g : G' \to \prod G_i$ which makes the required diagram commutative. It is the homomorphism such that $g(x')_i = g_i(x')$ for $x' \in G'$ and each $i \in I$.

Let A, B be objects of a category $\mathfrak{A}$. We note that the product of A, B is universal in the category whose objects consist of pairs of morphisms $f : C \to A$ and $g : C \to B$ in $\mathfrak{A}$, and whose morphisms are described as follows. Let $f' : C' \to A$ and $g' : C' \to B$ be another pair. Then a morphism from the first pair to the second is a morphism $h : C \to C'$ in $\mathfrak{A}$, making the following diagram commutative:



The situation is similar for the product of a family $\{A_i\}_{i \in I}$.

We shall also meet the dual notion: Let $\{A_i\}_{i \in I}$ be a family of objects in a category $\mathfrak{A}$. By their **coproduct** one means a pair $(S, \{f_i\}_{i \in I})$ consisting of an object S and a family of morphisms

$$\{f_i : A_i \to S\},$$

satisfying the following property. Given a family of morphisms $\{g_i : A_i \to C\}$, there exists a unique morphism $h : S \to C$ such that $h \circ f_i = g_i$ for all i.

In the product and coproduct, the morphism h will be said to be the morphism **induced** by the family $\{g_i\}$.

Examples. Let $\mathcal{S}$ be the category of sets. *Then coproducts exist.* For instance, let S, S' be sets. Let T be a set having the same cardinality as S' and disjoint from S. Let $f_1 : S \to S$ be the identity, and $f_2 : S' \to T$ be a bijection. Let U be the union of S and T. Then (U, f_1, f_2) is a coproduct for S, S', viewing f_1, f_2 as maps into U.

Let $\mathcal{S}_0$ be the category of pointed sets. Its objects consist of pairs (S, x) where S is a set and x is an element of S. A morphism of (S, x) into (S', x') in this category is a map $g : S \to S'$ such that $g(x) = x'$. *Then the coproduct of* (S, x) *and* (S', x') *exists in this category*, and can be constructed as follows. Let T be a set whose cardinality is the same as that of S', and such that $T \cap S = \{x\}$. Let $U = S \cup T$, and let

$$f_1 : (S, x) \to (U, x)$$

be the map which induces the identity on S. Let

$$f_2 : (S', x') \to (U, x)$$

be a map sending x' to x and inducing a bijection of $S' - \{x'\}$ on $T - \{x\}$. Then the triple $((U, x), f_1, f_2)$ is a coproduct for (S, x) and (S', x') in the category of pointed sets.

Similar constructions can be made for the coproduct of arbitrary families of sets or pointed sets. The category of pointed sets is especially important in homotopy theory.

Coproducts are universal objects. Indeed, let $\mathfrak{A}$ be a category, and let $\{A_i\}$ be a family of objects in $\mathfrak{A}$. We now define $\mathcal{C}$. We let objects of $\mathcal{C}$ be the families of morphisms $\{f_i : A_i \to B\}_{i \in I}$ and given two such families,

$$\{f_i : A_i \to B\} \quad \text{and} \quad \{f'_i : A_i \to B'\},$$

we define a morphism from the first into the second to be a morphism $\varphi : B \to B'$ in $\mathfrak{A}$ such that $\varphi \circ f_i = f'_i$ for all i. Then a coproduct of $\{A_i\}$ is simply a universal object in $\mathcal{C}$.

The coproduct of $\{A_i\}$ will be denoted by

$$\coprod_{i \in I} A_i.$$

The coproduct of two objects A, B will also be denoted by $A \amalg B$.

By the general uniqueness statement, we see that it is uniquely determined, up to a unique isomorphism.

Example. Let R be the category of commutative rings. Given two such rings A, B one may form the tensor product, and there are natural ring-homomorphisms $A \to A \otimes B$ and $B \to A \otimes B$ such that

$$a \mapsto a \otimes 1 \text{ and } b \mapsto 1 \otimes b \text{ for } a \in A \text{ and } b \in B.$$

Then the tensor product is a coproduct in the category of commutative rings.

Fiber products and coproducts
Pull-backs and push-outs

Let $\mathcal{C}$ be a category. Let Z be an object of $\mathcal{C}$. Then we have a new category, that of objects over Z, denoted by $\mathcal{C}_Z$. The objects of $\mathcal{C}_Z$ are morphisms:

$$f: X \to Z \text{ in } \mathcal{C}$$

A morphism from f to $g: Y \to Z$ in $\mathcal{C}_Z$ is merely a morphism $h: X \to Y$ in $\mathcal{C}$ which makes the following diagram commutative.

$$\begin{array}{ccc} X & \xrightarrow{h} & Y \\ & {\scriptstyle f}\searrow \quad \swarrow{\scriptstyle g} & \\ & Z & \end{array}$$

A **product** in $\mathcal{C}_Z$ is called the **fiber product** of f and g in $\mathcal{C}$ and is denoted by $X \times_Z Y$, together with its natural morphisms on X, Y over Z, which are sometimes not denoted by anything, but which we denote by p_1, p_2.

$$\begin{array}{ccc} & X \times_Z Y & \\ {\scriptstyle p_1}\swarrow & & \searrow{\scriptstyle p_2} \\ X & & Y \\ {\scriptstyle f}\searrow & & \swarrow{\scriptstyle g} \\ & Z & \end{array}$$

Fibered products and coproducts exist in the category of abelian groups

The fibered product of two homomorphisms $f: X \to Z$ and $g: Y \to Z$ is the subgroup of $X \times Y$ consisting of all pairs (x, y) such that

$$f(x) = g(y).$$

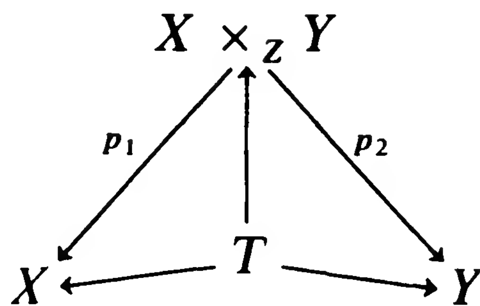
The coproduct of two homomorphisms $f: Z \to X$ and $g: Z \to Y$ is the factor group $(X \oplus Y)/W$ where W is the subgroup of $X \oplus Y$ consisting of all elements $(f(z), -g(z))$ with $z \in Z$.
We leave the simple verification to the reader (see Exercises 50–56).

In the fiber product diagram, one also calls p_1 the **pull-back** of g by f, and p_2 the **pull-back** of f by g. The fiber product satisfies the following universal mapping property:

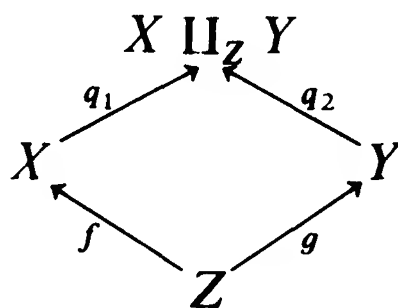
Given any object T in $\mathcal{C}$ and morphisms making the following diagram commutative:

$$\begin{array}{ccc} & T & \\ \swarrow & & \searrow \\ X & & Y \\ {\scriptstyle f}\searrow & & \swarrow{\scriptstyle g} \\ & Z & \end{array}$$

there exists a unique morphism $T \to X \times_Z Y$ *making the following diagram commutative:*



Dually, we have the notion of **coproduct** in the category of morphisms $f: Z \to X$ with a fixed object Z as the object of departure of the morphisms. This category could be denoted by $\mathcal{C}^Z$. We reverse the arrows in the preceding discussion. Given two objects f and $g: Z \to Y$ in this category, we have the notion of their coproduct. It is denoted by $X \amalg_Z Y$, with morphisms q_1, q_2, as in the following commutative diagram:



satisfying the dual universal property of the fiber product. We call it the **fibered coproduct**. We call q_1 the **push-out** of g by f, and q_2 the **push-out** of f by g.

Example. Let S be the category of sets. Given two maps f, g as above, their product is the set of all pairs $(x, y) \in X \times Y$ such that $f(x) = g(y)$.

Functors

Let $\mathcal{A}$, $\mathcal{B}$ be categories. A **covariant functor** F of $\mathcal{A}$ into $\mathcal{B}$ is a rule which to each object A in $\mathcal{A}$ associates an object $F(A)$ in $\mathcal{B}$, and to each morphism $f: A \to B$ associates a morphism $F(f): F(A) \to F(B)$ such that:

FUN 1. For all A in $\mathcal{A}$ we have $F(\mathrm{id}_A) = \mathrm{id}_{F(A)}$.

FUN 2. If $f: A \to B$ and $g: B \to C$ are two morphisms of $\mathcal{A}$ then

$$F(g \circ f) = F(g) \circ F(f).$$

Example. If to each group G we associate its set (stripped of the group structure) we obtain a functor from the category of groups into the category of sets, provided that we associate with each group-homomorphism itself, viewed only as a set-theoretic map. Such a functor is called a **stripping** functor or **forgetful functor**.

We observe that a functor transforms isomorphisms into isomorphisms, because $f \circ g = \mathrm{id}$ implies $F(f) \circ F(g) = \mathrm{id}$ also.

We can define the notion of a **contravariant functor** from $\mathcal{A}$ into $\mathcal{B}$ by using essentially the same definition, but reversing all arrows $F(f)$, i.e. to each morphism $f: A \to B$ the contravariant functor associates a morphism

$$F(f): F(B) \to F(A)$$

(going in the opposite direction), such that, if

$$f: A \to B \quad \text{and} \quad g: B \to C$$

are morphisms in $\mathfrak{A}$, then

$$F(g \circ f) = F(f) \circ F(g).$$

Sometimes a functor is denoted by writing f_* instead of $F(f)$ in the case of a covariant functor, and by writing f^* in the case of a contravariant functor.

Example. The association $S \mapsto F_{ab}(S)$ is a covariant functor from the category of sets to the category of abelian groups.

Example. The association which to each group associates its completion with respect to the family of subgroups of finite index is a functor from the category of groups to the category of groups.

Example. Let p be a prime number. Let $\mathfrak{C}$ be the category of p-divisible abelian groups. The association $A \mapsto T_p(A)$ is a covariant functor of $\mathfrak{C}$ into abelian groups (actually $\mathbf{Z}_p$-modules).

Example. Exercise 49 will show you an example of the group of automorphisms of a forgetful functor.

Example. Let **Man** be the category of compact manifolds. Then the homology is a covariant functor from **Man** into graded abelian groups. The cohomology is a contravariant functor into the category of graded algebras (over the ring of coefficients). The product is the cup product. If the cohomology is taken with coefficients in a field of characteristic 0 (for simplicity), then the cohomology commutes with products. Since cohomology is contravariant, this means that the cohomology of a product is the coproduct of the cohomology of the factors. It turns out that the coproduct is the tensor product, with the graded product, which also gives an example of the use of tensor products. See M. GREENBERG and J. HARPER, *Algebraic Topology* (Benjamin-Addison-Wesley), 1981, Chapter 29.

Example. Let $\mathfrak{C}$ be the category of pointed topological spaces (satisfying some mild conditions), i.e. pairs (X, x_0) consisting of a space X and a point x_0. In topology one defines the connected sum of such spaces (X, x_0) and (Y, y_0), glueing X, Y together at the selected point. This connected sum is a coproduct in the category of such pairs, where the morphisms are the continuous maps $f: X \to Y$ such that $f(x_0) = y_0$. Let π_1 denote the fundamental group. Then $(X, x_0) \mapsto \pi_1(X, x_0)$ is a covariant functor from $\mathfrak{C}$ into the category of groups, commuting with coproducts. (The existence of coproducts in the category of groups will be proved in §12.)

Example. Suppose we have a morphism $f: X \to Y$ in a category $\mathcal{C}$. By **a section** of f, one means a morphism $g: Y \to X$ such that $g \circ f = \text{id}$. Suppose there exists a covariant functor H from this category to groups such that $H(Y) = \{e\}$ and $H(X) \neq \{e\}$. Then there is no section of f. This is immediate from the formula $H(g \circ f) = \text{id}$, and $H(f) = $ trivial homomorphism. In topology one uses the homology functor to show, for instance, that the unit circle X is not a retract of the closed unit disc with respect to the inclusion mapping f. (Topologists use the word "retract" instead of "section".)

Example. Let $\mathfrak{A}$ be a category and A a fixed object in $\mathfrak{A}$. Then we obtain a covariant functor

$$M_A : \mathfrak{A} \to \mathfrak{S}$$

by letting $M_A(X) = \text{Mor}(A, X)$ for any object X of $\mathfrak{A}$. If $\varphi : X \to X'$ is a morphism, we let

$$M_A(\varphi) : \text{Mor}(A, X) \to \text{Mor}(A, X')$$

be the map given by the rule

$$g \mapsto \varphi \circ g$$

for any $g \in \text{Mor}(A, X)$,

$$A \xrightarrow{g} X \xrightarrow{\varphi} X'.$$

The axioms **FUN 1** and **FUN 2** are trivially verified.

Similarly, for each object B of $\mathfrak{A}$, we have a contravariant functor

$$M^B : \mathfrak{A} \to \mathfrak{S}$$

such that $M^B(Y) = \text{Mor}(Y, B)$. If $\psi : Y' \to Y$ is a morphism, then

$$M^B(\psi) : \text{Mor}(Y, B) \to \text{Mor}(Y', B)$$

is the map given by the rule

$$f \mapsto f \circ \psi$$

for any $f \in \text{Mor}(Y, B)$,

$$Y' \xrightarrow{\psi} Y \xrightarrow{f} B.$$

The preceding two functors are called the **representation functors**.

Example. Let $\mathfrak{A}$ be the category of abelian groups. Fix an abelian group A. The association $X \mapsto \text{Hom}(A, X)$ is a covariant functor from $\mathfrak{A}$ into itself. The association $X \mapsto \text{Hom}(X, A)$ is a contravariant functor of $\mathfrak{A}$ into itself.

Example. We assume you know about the tensor product. Let A be a commutative ring. Let M be an A-module. The association $X \mapsto M \otimes X$ is a covariant functor from the category of A-modules into itself.

Observe that products and coproducts were defined in a way compatible with the representation functor into the category of sets. Indeed, given a product P

of two objects A and B, then for every object X the set Mor(X, P) is a product of the sets Mor(X, A) and Mor(X, B) in the category of sets. This is merely a reformulation of the defining property of products in arbitrary categories. The system really works.

Let $\mathfrak{A}$, $\mathfrak{B}$ be two categories. The functors of $\mathfrak{A}$ into $\mathfrak{B}$ (say covariant, and in one variable) can be viewed as the objects of a category, whose morphisms are defined as follows. Let L, M be two such functors. A **morphism** $H: L \to M$ (also called a **natural transformation**) is a rule which to each object X of $\mathfrak{A}$ associates a morphism

$$H_X: L(X) \to M(X)$$

such that for any morphism $f: X \to Y$ the following diagram is commutative:

$$\begin{array}{ccc} L(X) & \xrightarrow{H_X} & M(X) \\ {\scriptstyle L(f)}\downarrow & & \downarrow{\scriptstyle M(f)} \\ L(Y) & \xrightarrow[H_Y]{} & M(Y) \end{array}$$

We can therefore speak of **isomorphisms** of functors. A functor is **representable** if it is isomorphic to a representation functor as above.

As Grothendieck pointed out, one can use the representation functor to transport the notions of certain structures on sets to arbitrary categories. For instance, let $\mathfrak{A}$ be a category and G an object of $\mathfrak{A}$. We say that G is a **group object** in $\mathfrak{A}$ if for each object X of $\mathfrak{A}$ we are given a group structure on the set Mor(X, G) in such a way that the association

$$X \mapsto \mathrm{Mor}(X, G)$$

is functorial (i.e. is a functor from $\mathfrak{A}$ into the category of groups). One sometimes denotes the set Mor(X, G) by $G(X)$, and thinks of it as the set of points of G in X. To justify this terminology, the reader is referred to Chapter IX, §2.

Example. Let **Var** be the category of projective non-singular varieties over the complex numbers. To each object X in **Var** one can associate various groups, e.g. Pic(X) (the group of divisor classes for rational equivalence), which is a contravariant functor into the category of abelian groups. Let $\mathrm{Pic}_0(X)$ be the subgroup of classes algebraically equivalent to 0. Then Pic_0 is representable.

In the fifties and sixties Grothendieck was the one who emphasized the importance of the representation functors, and the possibility of transposing to any category notions from more standard categories by means of the representation functors. He himself proved that a number of important functors in algebraic geometry are representable.

§12. FREE GROUPS

We now turn to the coproduct in the category of groups. First a remark. Let $G = \prod G_i$ be a direct product of groups.

We observe that each G_j admits an injective homomorphism into the product, on the j-th component, namely the map $\lambda_j : G_j \to \prod_i G_i$ such that for x in G_j, the i-th component of $\lambda_j(x)$ is the unit element of G_i if $i \neq j$, and is equal to x itself if $i = j$. This embedding will be called the **canonical** one. But we still don't have a coproduct of the family, because the factors commute with each other. To get a coproduct one has to work somewhat harder.

Let G be a group and S a subset of G. We recall that G is **generated** by S if every element of G can be written as a finite product of elements of S and their inverses (the empty product being always taken as the unit element of G). Elements of S are then called **generators**. If there exists a finite set of generators for G we call G **finitely generated**. If S is a set and $\varphi : S \to G$ is a map, we say that φ **generates** G if its image generates G.

Let S be a set, and $f : S \to F$ a map into a group. Let $g : S \to G$ be another map. If $f(S)$ (or as we also say, f) generates F, then it is obvious that there exists at most one homomorphism ψ of F into G which makes the following diagram commutative:

$$\begin{array}{ccc} S & \xrightarrow{f} & F \\ & {\scriptstyle g}\searrow \quad \swarrow{\scriptstyle \psi} & \\ & G & \end{array}$$

We now consider the category $\mathcal{C}$ whose objects are the maps of S into groups. If $f : S \to G$ and $f' : S \to G'$ are two objects in this category, we define a morphism from f to f' to be a homomorphism $\varphi : G \to G'$ such that $\varphi \circ f = f'$, i.e. the diagram is commutative:

$$\begin{array}{ccc} & & G \\ & \nearrow^{f} & \downarrow{\scriptstyle \varphi} \\ S & & \\ & \searrow_{f'} & G' \end{array}$$

By a **free group** determined by S, we shall mean a universal element in this category.

Proposition 12.1. *Let S be a set. Then there exists a free group (F, f) determined by S. Furthermore, f is injective, and F is generated by the image of f.*

Proof. (I owe this proof to J. Tits.) We begin with a lemma.

Lemma 12.2. *There exists a set I and a family of groups $\{G_i\}_{i \in I}$ such that, if $g\colon S \to G$ is a map of S into a group G, and g generates G, then G is isomorphic to some G_i.*

Proof. This is a simple exercise in cardinalities, which we carry out. If S is finite, then G is finite or denumerable. If S is infinite, then the cardinality of G is $\leqq$ the cardinality of S because G consists of finite products of elements of $g(S)$. Let T be a set which is infinite denumerable if S is finite, and has the same cardinality as S if S is infinite. For each non-empty subset H of T, let Γ_H be the set of group structures on H. For each $\gamma \in \Gamma_H$, let H_γ be the set H, together with the group structure γ. Then the family $\{H_\gamma\}$ for $\gamma \in \Gamma_H$ and H ranging over subsets of T is the desired family.

We return to the proof of the proposition. For each $i \in I$ we let M_i be the set of mappings of S into G_i. For each map $\varphi \in M_i$, we let $G_{i,\varphi}$ be the set-theoretic product of G_i and the set with one element $\{\varphi\}$, so that $G_{i,\varphi}$ is the "same" group as G_i indexed by φ. We let

$$F_0 = \prod_{i \in I} \prod_{\varphi \in M_i} G_{i,\varphi}$$

be the Cartesian product of the groups $G_{i,\varphi}$. We define a map

$$f_0 : S \to F_0$$

by sending S on the factor $G_{i,\varphi}$ by means of φ itself. We contend that given a map $g\colon S \to G$ of S into a group G, there exists a homomorphism $\psi_*\colon F_0 \to G$ making the usual diagram commutative:

$$\begin{array}{ccc} & & F_0 \\ & \overset{f_0}{\nearrow} & \\ S & & \downarrow \psi_* \\ & \underset{g}{\searrow} & \\ & & G \end{array}$$

That is, $\psi_* \circ f_0 = g$. To prove this, we may assume that g generates G, simply by restricting our attention to the subgroup of G generated by the image of g. By the lemma, there exists an isomorphism $\lambda : G \to G_i$ for some i, and $\lambda \circ g$ is an element ψ of M_i. We let $\pi_{i,\psi}$ be the projection on the (i, ψ) factor, and we let $\psi_* = \lambda^{-1} \circ \pi_{i,\psi}$. Then the map ψ_* makes the following diagram commutative.

$$\begin{array}{ccc} S & \xrightarrow{f_0} & F_0 \\ {\scriptstyle g}\downarrow & \overset{\psi_*}{\swarrow} & \downarrow{\scriptstyle \pi_{i,\psi}} \\ G & \xrightarrow[\lambda]{} & G_{i,\psi} \end{array}$$

We let F be the subgroup of F_0 generated by the image of f_0, and we let f simply be equal to f_0, viewed as a map of S into F. We let g_* be the restriction of ψ_* to F. In this way, we see at once that the map g_* is the unique one making

our diagram commutative, and thus that (F, f) is the required free group. Furthermore, it is clear that f is injective.

For each set S we select one free group determined by S, and denote it by $(F(S), f_S)$ or briefly by $F(S)$. It is generated by the image of f_S. One may view S as contained in $F(S)$, and the elements of S are called **free** generators of $F(S)$. If $g: S \to G$ is a map, we denote by $g_*: F(S) \to G$ the homomorphism realizing the universality of our free group $F(S)$.

If $\lambda: S \to S'$ is a map of one set into another, we let $F(\lambda): F(S) \to F(S')$ be the map $(f_{S'} \circ \lambda)_*$.

$$\begin{array}{ccc} S & \xrightarrow{f_S} & F(S) \\ \lambda \downarrow & \searrow & \downarrow \lambda_* = F(\lambda) \\ S' & \xrightarrow[f_{S'}]{} & F(S') \end{array}$$

Then we may regard F as a functor from the category of sets to the category of groups (the functorial properties are trivially verified, and will be left to the reader).

If λ is surjective, then $F(\lambda)$ is also surjective.

We again leave the proof to the reader.

If two sets S, S' have the same cardinality, then they are isomorphic in the category of sets (an isomorphism being in this case a bijection!), and hence $F(S)$ is isomorphic to $F(S')$. If S has n elements, we call $F(S)$ the **free group on n generators**.

Let G be a group, and let S be the same set as G (i.e. G viewed as a set, without group structure). We have the identity map $g: S \to G$, and hence a surjective homomorphism

$$g_*: F(S) \to G$$

which will be called **canonical**. Thus every group is a factor group of a free group.

One can also construct groups by what is called **generators and relations**. Let S be a set, and $F(S)$ the free group. We assume that $f: S \to F(S)$ is an inclusion. Let R be a set of elements of $F(S)$. Each element of R can be written as a finite product

$$\prod_{\nu=1}^{n} x_\nu$$

where each x_ν is an element of S or an inverse of an element of S. Let N be the smallest normal subgroup of $F(S)$ containing R, i.e. the intersection of all normal subgroups of $F(S)$ containing R. Then $F(S)/N$ will be called the group **determined by the generators S and the relations R**.

Example. One shows easily that the group determined by one generator a, and the relation $\{a^2\}$, has order 2.

The canonical homomorphism $\varphi: F(S) \to F(S)/N$ satisfies the universal mapping property for homomorphisms ψ of $F(S)$ into groups G such that $\psi(x) = e$ for all $x \in R$. In view of this, one sometimes calls the group $F(S)/N$ the group determined by the generators S, and the relations $x = e$ (for all $x \in R$). For instance, the group in the preceding example would be called the group determined by the generator a, and the relation $a^2 = e$.

Let G be a group generated by a finite number of elements, and satisfying the relation $x^2 = e$ for all $x \in G$. What does G look like? It is easy to show that G is commutative. Then one can view G as a vector space over $\mathbf{Z}/2\mathbf{Z}$, so G is determined by its cardinality, up to isomorphism.

In Exercises 34 and 35, you will prove that there exist certain groups satisfying certain relations and with a given order, so that the group presented with these generators and relations can be completely determined. *A priori*, it is not even clear if a group given by generators and relations is finite. Even if it is finite, one does not know its order *a priori*. To show that a group of certain order exists, one has to use various means, a common means being to represent the group as a group of automorphisms of some object, for instance the symmetries of a geometric object. This will be the method suggested for the groups in Exercises 34 and 35, mentioned above.

Example. Let G be a group. For $x, y \in G$ define $[x, y] = xyx^{-1}y^{-1}$ (the commutator) and ${}^xy = xyx^{-1}$ (the conjugate). Then one has the cocycle relation

$$[x, yz] = [x, y]\,{}^y[x, z].$$

Furthermore, suppose $x, y, z \in G$ and

$$[x, y] = y, \quad [y, z] = z, \quad [z, x] = x.$$

Then $x = y = z = e$. It is an exercise to prove these assertions, but one sees that certain relations imply that a group generated by x, y, z subject to those relations is necessarily trivial.

Next we give a somewhat more sophisticated example. We assume that the reader knows the basic terminology of fields and matrices as in Chapter XIII, but applied only to 2×2 matrices. Thus $SL_2(F)$ denotes the group of 2×2 matrices with components in a field F and determinant equal to 1.

Example. $SL_2(F)$. Let F be a field. For $b \in F$ and $a \in F$, $a \neq 0$, we let

$$u(b) = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}, \quad s(a) = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}, \quad \text{and } w = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Then it is immediately verified that:

SL 0. $s(a) = wu(a^{-1})wu(a)wu(a^{-1})$.

SL 1. u is an additive homomorphism.

SL 2. s is a multiplicative homomorphism.

SL 3. $w^2 = s(-1)$.

SL 4. $s(a)u(b)s(a^{-1}) = u(ba^2)$.

Now, conversely, suppose that G is an arbitrary group with generators $u(b)$ ($b \in F$) and w, such that if we define $s(a)$ for $a \neq 0$ by **SL 0**, then the relations **SL 1** through **SL 4** are satisfied. Then **SL 3** and **SL 4** show that $s(-1)$ is in the center, and $w^4 = e$. In addition, one verifies that:

SL 5. $ws(a) = s(a^{-1})w$.

Furthermore, one has the theorem:

Let G be the free group with generators $u(b)$, w and relations **SL 1** *through* **SL 4**, *defining $s(a)$ as in* **SL 0**. *Then the natural homomorphism*

$$G \to SL_2(F)$$

is an isomorphism.

Proofs of all the above statements will be found in my $\mathbf{SL_2(R)}$, Springer Verlag, reprint of Addison-Wesley, 1975, Chapter XI, §2. It takes about a page to carry out the proof.

If $F = \mathbf{Q}_p$ is the field of p-adic numbers, then Ihara [Ih 66] proved that every discrete torsion free subgroup of $SL_2(\mathbf{Q}_p)$ is free. Serre put this theorem in the context of a general theory concerning groups acting on trees [Se 80].

[Ih 66] Y. IHARA, On discrete subgroups of the two by two projective linear group over p-adic fields, *J. Math. Soc. Japan* **18** (1966) pp. 219–235

[Se 80] J.-P. SERRE, *Trees*, Springer Verlag 1980

Further examples. For further examples of free group constructions, see Exercises 54 and 56. For examples of free groups occurring (possibly conjecturally) in Galois theory, see Chapter VI, §2, Example 9, and the end of Chapter VI, §14.

Proposition 12.3. *Coproducts exist in the category of groups.*

Proof. Let $\{G_i\}_{i \in I}$ be a family of groups. We let $\mathcal{C}$ be the category whose objects are families of group-homomorphisms

$$\{g_i : G_i \to G\}_{i \in I}$$

and whose morphisms are the obvious ones. We must find a universal element in this category. For each index i, we let S_i be the same set as G_i if G_i is infinite, and we let S_i be denumerable if G_i is finite. We let S be a set having the same cardinality as the set-theoretic disjoint union of the sets S_i (i.e. their coproduct in the category of sets). We let Γ be the set of group structures on S, and for each $\gamma \in \Gamma$, we let Φ_γ be the set of all families of homomorphisms

$$\varphi = \{\varphi_i : G_i \to S_\gamma\}.$$

Each pair (S_γ, φ), where $\varphi \in \Phi_\gamma$, is then a group, using φ merely as an index. We let

$$F_0 = \prod_{\gamma \in \Gamma} \prod_{\varphi \in \Phi_\gamma} (S_\gamma, \varphi),$$

and for each i, we define a homomorphism $f_i : G_i \to F_0$ by prescribing the component of f_i on each factor (S_γ, φ) to be the same as that of φ_i.

Let now $g = \{g_i : G_i \to G\}$ be a family of homomorphisms. Replacing G if necessary by the subgroup generated by the images of the g_i, we see that $\operatorname{card}(G) \leqq \operatorname{card}(S)$, because each element of G is a *finite* product of elements in these images. Embedding G as a factor in a product $G \times S_\gamma$ for some γ, we may assume that $\operatorname{card}(G) = \operatorname{card}(S)$. There exists a homomorphism $g_* : F_0 \to G$ such that

$$g_* \circ f_i = g_i$$

for all i. Indeed, we may assume without loss of generality that $G = S_\gamma$ for some γ and that $g = \psi$ for some $\psi \in \Phi_\gamma$. We let g_* be the projection of F_0 on the factor (S_γ, ψ).

Let F be the subgroup of F_0 generated by the union of the images of the maps f_i for all i. The restriction of g_* to F is the unique homomorphism satisfying $f_i \circ g_* = g_i$ for all i, and we have thus constructed our universal object.

Example. Let G_2 be a cyclic group of order 2 and let G_3 be a cyclic group of order 3. What is the coproduct? The answer is neat. It can be shown that $G_2 \amalg G_3$ is the group generated by two elements S, T with relations $S^2 = 1$, $(ST)^3 = 1$. The groups G_2 and G_3 are embedded in $G_2 \amalg G_3$ by sending G_2 on the cyclic group generated by S and sending G_3 on the cyclic group generated by ST. This is done by representing the group as follows. Let

$$G = SL_2(\mathbf{Z})/\pm 1.$$

As we have seen in an example of §5, the group G operates on the upper half-plane $\mathfrak{H}$. Let S, T be the maps given by

$$S(z) = -1/z \quad \text{and} \quad T(z) = z + 1.$$

Thus S and T are represented by the matrices

$$S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

and satisfy the relations $S^2 = 1$, $(ST)^3 = 1$. Readers will find a proof of several properties of S, T in Serre's *Course in Arithmetic* (Springer Verlag, 1973, Chapter VII, §1), including the fact that S, T generate G. It is an exercise from there to show that G is the coproduct of G_2 and G_3 as asserted.

Observe that these procedures go directly from the universal definition and construction in the proofs of Proposition 12.1 and Proposition 12.3 to the more explicit representation of the free group or the coproduct as the case may be. One relies on the following proposition.

Proposition 12.4. *Let G be a group and $\{G_i\}_{i \in I}$ a family of subgroups. Assume:*

(a) *The family generates G.*

(b) *If*

$$x = x_{i_1} \cdots x_{i_n} \quad \textit{with } x_{i_\nu} \in G_{i_\nu},\ x_{i_\nu} \neq e \textit{ and } i_\nu \neq i_{\nu+1} \textit{ for all } \nu,$$

then $x \neq e$.

Then the natural homomorphism of the coproduct of the family into G sending G_i on itself by the identity mapping is an isomorphism. In other words, simply put, G is the coproduct of the family of subgroups.

Proof. The homomorphism from the coproduct into G is surjective by the assumption that the family generates G. Suppose an element is in the kernel. Then such an element has a representation

$$x_{i_1} \cdots x_{i_n}$$

as in (b), mapping to the identity in G, so all $x_{i_\nu} = e$ and the element itself is equal to e, whence the homomorphism from the coproduct into G is injective, thereby proving the proposition.

Exercises 54 and 56 mentioned above give one illustration of the way Proposition 12.4 can be used. We now show another way, which we carry out for two subgroups. I am indebted to Eilenberg for the neat arrangement of the proof of the next proposition.

Proposition 12.5. *Let A, B be two groups whose set-theoretic intersection is $\{1\}$. There exists a group $A \circ B$ containing A, B as subgroups, such that $A \cap B = \{1\}$, and having the following property. Every element $\neq 1$ of $A \circ B$ has a unique expression as a product*

$$a_1 \cdots a_n \qquad (n \geqq 1, a_i \neq 1 \text{ all } i)$$

with $a_i \in A$ or $a_i \in B$, and such that if $a_i \in A$ then $a_{i+1} \in B$ and if $a_i \in B$ then $a_{i+1} \in A$.

Proof. Let $A \circ B$ be the set of sequences

$$a = (a_1, \ldots, a_n) \qquad (n \geqq 0)$$

such that either $n = 0$, and the sequence is empty or $n \geqq 1$, and then elements in the sequence belong to A or B, are $\neq 1$, and two consecutive elements of the sequence do not belong both to A or both to B. If $b = (b_1, \ldots, b_m)$, we define the product ab to be the sequence

$(a_1, \ldots, a_n, b_1, \ldots, b_m)$
if $a_n \in A, b_1 \in B$ or $a_n \in B, b_1 \in A$,

$(a_1, \ldots, a_n b_1, \ldots, b_m)$
if $a_n, b_1 \in A$ or $a_n, b_1 \in B$, and $a_n b_1 \neq 1$,

$(a_1, \ldots, a_{n-1})(b_2, \ldots, b_m)$ by induction,
if $a_n, b_1 \in A$ or $a_n, b_1 \in B$ and $a_n b_1 = 1$.

The case when $n = 0$ or $m = 0$ is included in the first case, and the empty sequence is the unit element of $A \circ B$. Clearly,

$$(a_1, \ldots, a_n)(a_n^{-1}, \ldots, a_1^{-1}) = \text{unit element},$$

so only associativity need be proved. Let $c = (c_1, \ldots, c_r)$.

First consider the case $m = 0$, i.e. b is empty. Then clearly $(ab)c = a(bc)$ and similarly if $n = 0$ or $r = 0$. Next consider the case $m = 1$. Let $b = (x)$ with $x \in A$, $x \neq 1$. We then verify in each possible case that $(ab)c = a(bc)$. These cases are as follows:

$(a_1, \ldots, a_n, x, c_1, \ldots, c_r)$ if $a_n \in B$ and $c_1 \in B$,

$(a_1, \ldots, a_n x, c_1, \ldots, c_r)$ if $a_n \in A, a_n x \neq 1, c_1 \in B$,

$(a_1, \ldots, a_n, x c_1, \ldots, c_r)$ if $a_n \in B, c_1 \in A, x c_1 \neq 1$,

$(a_1, \ldots, a_{n-1})(c_1, \ldots, c_r)$ if $a_n = x^{-1}$ and $c_1 \in B$,

$$(a_1, \dots, a_n)(c_2, \dots, c_r) \quad \text{if} \quad a_n \in B \quad \text{and} \quad c_1 = x^{-1},$$

$$(a_1, \dots, a_{n-1}, a_n x c_1, c_2, \dots, c_r) \quad \text{if} \quad a_n, c_1 \in A, a_n x c_1 \neq 1,$$

$$(a_1, \dots, a_{n-1})(c_2, \dots, c_r) \quad \text{if} \quad a_n, c_1 \in A \quad \text{and} \quad a_n x c_1 = 1.$$

If $m > 1$, then we proceed by induction. Write $b = b'b''$ with b' and b'' shorter. Then

$$(ab)c = (a(b'b''))c = ((ab')b'')c = (ab')(b''c),$$

$$a(bc) = a((b'b'')c) = a(b'(b''c)) = (ab')(b''c)$$

as was to be shown.

We have obvious injections of A and B into $A \circ B$, and identifying A, B with their images in $A \circ B$ we obtain a proof of our proposition.

We can prove the similar result for several factors. In particular, we get the following corollary for the free group.

Corollary 12.6. *Let $F(S)$ be the free group on a set S, and let $x_1, \dots, x_n$ be distinct elements of S. Let $\nu_1, \dots, \nu_r$ be integers $\neq 0$ and let $i_1, \dots, i_r$ be integers,*

$$1 \leqq i_1, \dots, i_r \leqq n$$

such that $i_j \neq i_{j+1}$ for $j = 1, \dots, r-1$. Then

$$x_{i_1}^{\nu_1} \cdots x_{i_r}^{\nu_r} \neq 1.$$

Proof. Let $G_1, \dots, G_n$ be the cyclic groups generated by $x_1, \dots, x_n$. Let $G = G_1 \circ \cdots \circ G_n$. Let

$$F(S) \to G$$

be the homomorphism sending each x_i on x_i, and all other elements of S on the unit element of G. Our assertion follows at once.

Corollary 12.7. *Let S be a set with n elements $x_1, \dots, x_n$, $n \geqq 1$. Let $G_1, \dots, G_n$ be the infinite cyclic groups generated by these elements. Then the map*

$$F(S) \to G_1 \circ \cdots \circ G_n$$

sending each x_i on itself is an isomorphism.

Proof. It is obviously surjective and injective.

Corollary 12.8. *Let $G_1, \dots, G_n$ be groups with $G_i \cap G_j = \{1\}$ if $i \neq j$. The homomorphism*

$$G_1 \amalg \cdots \amalg G_n \to G_1 \circ \cdots \circ G_n$$

of their coproduct into $G_1 \circ \cdots \circ G_n$ induced by the natural inclusion $G_i \to G_1 \circ \cdots \circ G_n$ is an isomorphism.

Proof. Again, it is obviously injective and surjective.

EXERCISES

1. Show that every group of order $\leqq 5$ is abelian.
2. Show that there are two non-isomorphic groups of order 4, namely the cyclic one, and the product of two cyclic groups of order 2.
3. Let G be a group. A **commutator** in G is an element of the form $aba^{-1}b^{-1}$ with a, $b \in G$. Let G^c be the subgroup generated by the commutators. Then G^c is called the **commutator subgroup**. Show that G^c is normal. Show that any homomorphism of G into an abelian group factors through G/G^c.
4. Let H, K be subgroups of a finite group G with $K \subset N_H$. Show that

$$\#(HK) = \frac{\#(H)\#(K)}{\#(H \cap K)}.$$

5. **Goursat's Lemma.** Let G, G' be groups, and let H be a subgroup of $G \times G'$ such that the two projections $p_1 : H \to G$ and $p_2 : H \to G'$ are surjective. Let N be the kernel of p_2 and N' be the kernel of p_1. One can identify N as a normal subgroup of G, and N' as a normal subgroup of G'. Show that the image of H in $G/N \times G'/N'$ is the graph of an isomorphism

$$G/N \approx G'/N'.$$

6. Prove that the group of inner automorphisms of a group G is normal in Aut(G).
7. Let G be a group such that Aut(G) is cyclic. Prove that G is abelian.
8. Let G be a group and let H, H' be subgroups. By a **double coset** of H, H' one means a subset of G of the form HxH'.
   (a) Show that G is a disjoint union of double cosets.
   (b) Let $\{c\}$ be a family of representatives for the double cosets. For each $a \in G$ denote by $[a]H'$ the conjugate $aH'a^{-1}$ of H'. For each c we have a decomposition into ordinary cosets

$$H = \bigcup_c x_c(H \cap [c]H'),$$

   where $\{x_c\}$ is a family of elements of H, depending on c. Show that the elements $\{x_c c\}$ form a family of left coset representatives for H' in G; that is,

$$G = \bigcup_{x_c} \bigcup_{x_c} x_c cH',$$

   and the union is disjoint. (Double cosets will not emerge further until Chapter XVIII.)
9. (a) Let G be a group and H a subgroup of finite index. Show that there exists a normal subgroup N of G contained in H and also of finite index. [*Hint*: If $(G : H) = n$, find a homomorphism of G into S_n whose kernel is contained in H.]
   (b) Let G be a group and let H_1, H_2 be subgroups of finite index. Prove that $H_1 \cap H_2$ has finite index.
10. Let G be a group and let H be a subgroup of finite index. Prove that there is only a finite number of right cosets of H, and that the number of right cosets is equal to the number of left cosets.

11. Let G be a group, and A a normal abelian subgroup. Show that G/A operates on A by conjugation, and in this manner get a homomorphism of G/A into $\mathrm{Aut}(A)$.

Semidirect product

12. Let G be a group and let H, N be subgroups with N normal. Let γ_x be conjugation by an element $x \in G$.
    (a) Show that $x \mapsto \gamma_x$ induces a homomorphism $f: H \mapsto \mathrm{Aut}(N)$.
    (b) If $H \cap N = \{e\}$, show that the map $H \times N \to HN$ given by $(x, y) \mapsto xy$ is a bijection, and that this map is an isomorphism if and only if f is trivial, i.e. $f(x) = \mathrm{id}_N$ for all $x \in H$.

    We define G to be the **semidirect product** of H and N if $G = NH$ and $H \cap N = \{e\}$.
    (c) Conversely, let N, H be groups, and let $\psi: H \to \mathrm{Aut}(N)$ be a given homomorphism. Construct a semidirect product as follows. Let G be the set of pairs (x, h) with $x \in N$ and $h \in H$. Define the composition law

$$(x_1, h_1)(x_2, h_2) = (x_1{}^{\psi(h_1)}x_2, h_1h_2).$$

    Show that this is a group law, and yields a semidirect product of N and H, identifying N with the set of elements $(x, 1)$ and H with the set of elements $(1, h)$.

13. (a) Let H, N be normal subgroups of a finite group G. Assume that the orders of H, N are relatively prime. Prove that $xy = yx$ for all $x \in H$ and $y \in N$, and that $H \times N \approx HN$.
    (b) Let $H_1, \ldots, H_r$ be normal subgroups of G such that the order of H_i is relatively prime to the order of H_j for $i \neq j$. Prove that

$$H_1 \times \ldots \times H_r \approx H_1 \cdots H_r.$$

    Example. If the Sylow subgroups of a finite group are normal, then G is the direct product of its Sylow subgroups.

14. Let G be a finite group and let N be a normal subgroup such that N and G/N have relatively prime orders.
    (a) Let H be a subgroup of G having the same order as G/N. Prove that $G = HN$.
    (b) Let g be an automorphism of G. Prove that $g(N) = N$.

Some operations

15. Let G be a finite group operating on a finite set S with $\#(S) \geqq 2$. Assume that there is only one orbit. Prove that there exists an element $x \in G$ which has no fixed point, i.e. $xs \neq s$ for all $s \in S$.

16. Let H be a proper subgroup of a finite group G. Show that G is not the union of all the conjugates of H. (But see Exercise 23 of Chapter XIII.)

17. Let X, Y be finite sets and let C be a subset of $X \times Y$. For $x \in X$ let $\varphi(x)$ = number of elements $y \in Y$ such that $(x, y) \in C$. Verify that

$$\#(C) = \sum_{x \in X} \varphi(x).$$

Remark. A subset C as in the above exercise is often called a **correspondence**, and $\varphi(x)$ is the number of elements in Y which correspond to a given element $x \in X$.

18. Let S, T be finite sets. Show that $\#\text{Map}(S, T) = (\#T)^{\#(S)}$.

19. Let G be a finite group operating on a finite set S.
    (a) For each $s \in S$ show that

$$\sum_{t \in Gs} \frac{1}{\#(Gt)} = 1.$$

    (b) For each $x \in G$ define $f(x)$ = number of elements $s \in S$ such that $xs = s$. Prove that the number of orbits of G in S is equal to

$$\frac{1}{\#(G)} \sum_{x \in G} f(x).$$

Throughout, p is a prime number.

20. Let P be a p-group. Let A be a normal subgroup of order p. Prove that A is contained in the center of P.

21. Let G be a finite group and H a subgroup. Let P_H be a p-Sylow subgroup of H. Prove that there exists a p-Sylow subgroup P of G such that $P_H = P \cap H$.

22. Let H be a normal subgroup of a finite group G and assume that $\#(H) = p$. Prove that H is contained in every p-Sylow subgroup of G.

23. Let P, P' be p-Sylow subgroups of a finite group G.
    (a) If $P' \subset N(P)$ (normalizer of P), then $P' = P$.
    (b) If $N(P') = N(P)$, then $P' = P$.
    (c) We have $N(N(P)) = N(P)$.

Explicit determination of groups

24. Let p be a prime number. Show that a group of order p^2 is abelian, and that there are only two such groups up to isomorphism.

25. Let G be a group of order p^3, where p is prime, and G is not abelian. Let Z be its center. Let C be a cyclic group of order p.
    (a) Show that $Z \approx C$ and $G/Z \approx C \times C$.
    (b) Every subgroup of G of order p^2 contains Z and is normal.
    (c) Suppose $x^p = 1$ for all $x \in G$. Show that G contains a normal subgroup $H \approx C \times C$.

26. (a) Let G be a group of order pq, where p, q are primes and $p < q$. Assume that $q \not\equiv 1 \bmod p$. Prove that G is cyclic.
    (b) Show that every group of order 15 is cyclic.

27. Show that every group of order < 60 is solvable.

28. Let p, q be distinct primes. Prove that a group of order p^2q is solvable, and that one of its Sylow subgroups is normal.

29. Let p, q be odd primes. Prove that a group of order $2pq$ is solvable.

30. (a) Prove that one of the Sylow subgroups of a group of order 40 is normal.
(b) Prove that one of the Sylow subgroups of a group of order 12 is normal.

31. Determine all groups of order $\leqq 10$ up to isomorphism. In particular, show that a non-abelian group of order 6 is isomorphic to S_3.

32. Let S_n be the permutation group on n elements. Determine the p-Sylow subgroups of S_3, S_4, S_5 for $p = 2$ and $p = 3$.

33. Let σ be a permutation of a finite set I having n elements. Define $e(\sigma)$ to be $(-1)^m$ where

$$m = n - \text{number of orbits of } \sigma.$$

If $I_1, \dots, I_r$ are the orbits of σ, then m is also equal to the sum

$$m = \sum_{\nu=1}^{r} [\operatorname{card}(I_\nu) - 1].$$

If τ is a transposition, show that $e(\sigma\tau) = -e(\sigma)$ be considering the two cases when i, j lie in the same orbit of σ, or lie in different orbits. In the first case, $\sigma\tau$ has one more orbit and in the second case one less orbit than σ. In particular, the sign of a transposition is -1. Prove that $e(\sigma) = \varepsilon(\sigma)$ is the sign of the permutation.

34. (a) Let n be an even positive integer. Show that there exists a group of order $2n$, generated by two elements σ, τ such that $\sigma^n = e = \tau^2$, and $\sigma\tau = \tau\sigma^{n-1}$. (Draw a picture of a regular n-gon, number the vertices, and use the picture as an inspiration to get σ, τ.) This group is called the **dihedral group**.
(b) Let n be an odd positive integer. Let D_{4n} be the group generated by the matrices

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} \zeta & 0 \\ 0 & \zeta^{-1} \end{pmatrix}$$

where ζ is a primitive n-th root of unity. Show that D_{4n} has order $4n$, and give the commutation relations between the above generators.

35. Show that there are exactly two non-isomorphic non-abelian groups of order 8. (One of them is given by generators σ, τ with the relations

$$\sigma^4 = 1, \qquad \tau^2 = 1, \qquad \tau\sigma\tau = \sigma^3.$$

The other is the quaternion group.)

36. Let $\sigma = [123 \cdots n]$ in S_n. Show that the conjugacy class of σ has $(n-1)!$ elements. Show that the centralizer of σ is the cyclic group generated by σ.

37. (a) Let $\sigma = [i_1 \cdots i_m]$ be a cycle. Let $\gamma \in S_n$. Show that $\gamma\sigma\gamma^{-1}$ is the cycle $[\gamma(i_1) \cdots \gamma(i_m)]$.
(b) Suppose that a permutation σ in S_n can be written as a product of r disjoint cycles, and let $d_1, \dots, d_r$ be the number of elements in each cycle, in increasing order. Let τ be another permutation which can be written as a product of disjoint cycles, whose cardinalities are $d'_1, \dots, d'_s$ in increasing order. Prove that σ is conjugate to τ in S_n if and only if $r = s$ and $d_i = d'_i$ for all $i = 1, \dots, r$.

38. (a) Show that S_n is generated by the transpositions $[12], [13], \dots, [1n]$.
(b) Show that S_n is generated by the transpositions $[12], [23], [34], \dots, [n-1, n]$.

(c) Show that S_n is generated by the cycles $[12]$ and $[123 \ldots n]$.
(d) Assume that n is prime. Let $\sigma = [123 \ldots n]$ and let $\tau = [rs]$ be any transposition. Show that σ, τ generate S_n.

Let G be a finite group operating on a set S. Then G operates in a natural way on the Cartesian product $S^{(n)}$ for each positive integer n. We define the operation on S to be **n-transitive** if given n distinct elements $(s_1, \ldots, s_n)$ and n distinct elements $(s'_1, \ldots, s'_n)$ of S, there exists $\sigma \in G$ such that $\sigma s_i = s'_i$ for all $i = 1, \ldots, n$.

39. Show that the action of the alternating group A_n on $\{1, \ldots, n\}$ is $(n-2)$-transitive.

40. Let A_n be the alternating group of even permutations of $\{1, \ldots, n\}$. For $j = 1, \ldots, n$ let H_j be the subgroup of A_n fixing j, so $H_j \approx A_{n-1}$, and $(A_n : H_j) = n$ for $n \geqq 3$. Let $n \geqq 3$ and let H be a subgroup of index n in A_n.
    (a) Show that the action of A_n on cosets of H by left translation gives an isomorphism A_n with the alternating group of permutations of A_n/H.
    (b) Show that there exists an automorphism of A_n mapping H_1 on H, and that such an automorphism is induced by an inner automorphism of S_n if and only if $H = H_i$ for some i.

41. Let H be a simple group of order 60.
    (a) Show that the action of H by conjugation on the set of its Sylow subgroups gives an imbedding $H \hookrightarrow A_6$.
    (b) Using the preceding exercise, show that $H \approx A_5$.
    (c) Show that A_6 has an automorphism which is not induced by an inner automorphism of S_6.

Abelian groups

42. Viewing $\mathbf{Z}, \mathbf{Q}$ as additive groups, show that $\mathbf{Q}/\mathbf{Z}$ is a torsion group, which has one and only one subgroup of order n for each integer $n \geqq 1$, and that this subgroup is cyclic.

43. Let H be a subgroup of a finite abelian group G. Show that G has a subgroup that is isomorphic to G/H.

44. Let $f: A \to A'$ be a homomorphism of abelian groups. Let B be a subgroup of A. Denote by A^f and A_f the image and kernel of f in A respectively, and similarly for B^f and B_f. Show that $(A:B) = (A^f : B^f)(A_f : B_f)$, in the sense that if two of these three indices are finite, so is the third, and the stated equality holds.

45. Let G be a finite cyclic group of order n, generated by an element σ. Assume that G operates on an abelian group A, and let $f, g: A \to A$ be the endomorphisms of A given by

$$f(x) = \sigma x - x \quad \text{and} \quad g(x) = x + \sigma x + \cdots + \sigma^{n-1}x.$$

Define the **Herbrand quotient** by the expression $q(A) = (A_f : A^g)/(A_g : A^f)$, provided both indices are finite. Assume now that B is a subgroup of A such that $GB \subset B$.
    (a) Define in a natural way an operation of G on A/B.
    (b) Prove that

$$q(A) = q(B)q(A/B)$$

in the sense that if two of these quotients are finite, so is the third, and the stated equality holds.
    (c) If A is finite, show that $q(A) = 1$.

(This exercise is a special case of the general theory of Euler characteristics discussed in Chapter XX, Theorem 3.1. After reading this, the present exercise becomes trivial. Why?)

Primitive groups

46. Let G operate on a set S. Let $S = \bigcup S_i$ be a partition of S into disjoint subsets. We say that the partition is **stable** under G if G maps each S_i onto S_j for some j, and hence G induces a permutation of the sets of the partition among themselves. There are two partitions of S which are obviously stable: the partition consisting of S itself, and the partition consisting of the subsets with one element. Assume that G operates transitively, and that S has more than one element. Prove that the following two conditions are equivalent:

    PRIM 1. The only partitions of S which are stable are the two partitions mentioned above.

    PRIM 2. If H is the isotropy group of an element of S, then H is a maximal subgroup of G.

    These two conditions define what is known as a **primitive group**, or more accurately, a **primitive operation** of G on S.

    Instead of saying that the operation of a group G is 2-transitive, one also says that it is **doubly transitive**.

47. Let a finite group G operate transitively and faithfully on a set S with at least 2 elements and let H be the isotropy group of some element s of S. (All the other isotropy groups are conjugates of H.) Prove the following:
    (a) G is doubly transitive if and only if H acts transitively on the complement of s in S.
    (b) G is doubly transitive if and only if $G = HTH$, where T is a subgroup of G of order 2 not contained in H.
    (c) If G is doubly transitive, and $(G : H) = n$, then

    $$\#(G) = d(n - 1)n,$$

    where d is the order of the subgroup fixing two elements. Furthermore, H is a maximal subgroup of G, i.e. G is primitive.

48. Let G be a group acting transitively on a set S with at least 2 elements. For each $x \in G$ let $f(x)$ = number of elements of S fixed by x. Prove:

    (a) $\sum_{x \in G} f(x) = \#(G)$.
    (b) G is doubly transitive if and only if

    $$\sum_{x \in G} f(x)^2 = 2\ \#(G).$$

49. **A group as an automorphism group.** Let G be a group and let **Set**(G) be the category of G-sets (i.e. sets with a G-operation). Let F: **Set**$(G) \to$ **Set** be the forgetful functor, which to each G-set assigns the set itself. Show that $\mathrm{Aut}(F)$ is naturally isomorphic to G.

Fiber products and coproducts
Pull-backs and push-outs

50. (a) Show that fiber products exist in the category of abelian groups. In fact, if X, Y are abelian groups with homomorphisms $f: X \to Z$ and $g: Y \to Z$ show that $X \times_Z Y$ is the set of all pairs (x, y) with $x \in X$ and $y \in Y$ such that $f(x) = g(y)$. The maps p_1, p_2 are the projections on the first and second factor respectively.
(b) Show that the pull-back of a surjective homomorphism is surjective.

51. (a) Show that fiber products exist in the category of sets.
(b) In any category $\mathcal{C}$, consider the category $\mathcal{C}_Z$ of objects over Z. Let $h: T \to Z$ be a fixed object in this category. Let F be the functor such that

$$F(X) = \mathrm{Mor}_Z(T, X),$$

where X is an object over Z, and Mor_Z denotes morphisms over Z. Show that F transforms fiber products over Z into fiber products in the category of sets. (Actually, once you have understood the definitions, this is tautological.)

52. (a) Show that push-outs (i.e. fiber coproducts) exist in the category of abelian groups. In this case the fiber coproduct of two homomorphisms f, g as above is denoted by $X \oplus_Z Y$. Show that it is the factor group

$$X \oplus_Z Y = (X \oplus Y)/W,$$

where W is the subgroup consisting of all elements $(f(z), -g(z))$ with $z \in Z$.
(b) Show that the push-out of an injective homomorphism is injective.

Remark. After you have read about modules over rings, you should note that the above two exercises apply to modules as well as to abelian groups.

53. Let H, G, G' be groups, and let

$$f: H \to G, \qquad g: H \to G'$$

be two homomorphisms. Define the notion of coproduct of these two homomorphisms over H, and show that it exists.

54. (Tits). Let G be a group and let $\{G_i\}_{i \in I}$ be a family of subgroups generating G. Suppose G operates on a set S. For each $i \in I$, suppose given a subset S_i of S, and let s be a point of $S - \bigcup_i S_i$. Assume that for each $g \in G_i - \{e\}$, we have

$$gS_j \subset S_i \text{ for all } j \neq i, \quad \text{and} \quad g(s) \in S_i \text{ for all } i.$$

Prove that G is the coproduct of the family $\{G_i\}_{i \in I}$. (*Hint*: Suppose a product $g_1 \cdots g_m = \mathrm{id}$ on S. Apply this product to s, and use Proposition 12.4.)

55. Let $M \in GL_2(\mathbf{C})$ (2×2 complex matrices with non-zero determinant). We let

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \text{ and for } z \in \mathbf{C} \text{ we let } M(z) = \frac{az + b}{cz + d}.$$

If $z = -d/c$ ($c \neq 0$) then we put $M(z) = \infty$. Then you can verify (and you should have seen something like this in a course in complex analysis) that $GL_2(\mathbf{C})$ thus operates on $\mathbf{C} \cup \{\infty\}$. Let λ, λ' be the eigenvalues of M viewed as a linear map on $\mathbf{C}^2$. Let W, W' be the corresponding eigenvectors,

$$W = {}^t(w_1, w_2) \text{ and } W' = {}^t(w_1', w_2').$$

By a **fixed point** of M on $\mathbf{C}$ we mean a complex number z such that $M(z) = z$. *Assume that M has two distinct fixed points* $\neq \infty$.

(a) Show that there cannot be more than two fixed points and that these fixed points are $w = w_1/w_2$ and $w' = w'_1/w'_2$. In fact one may take

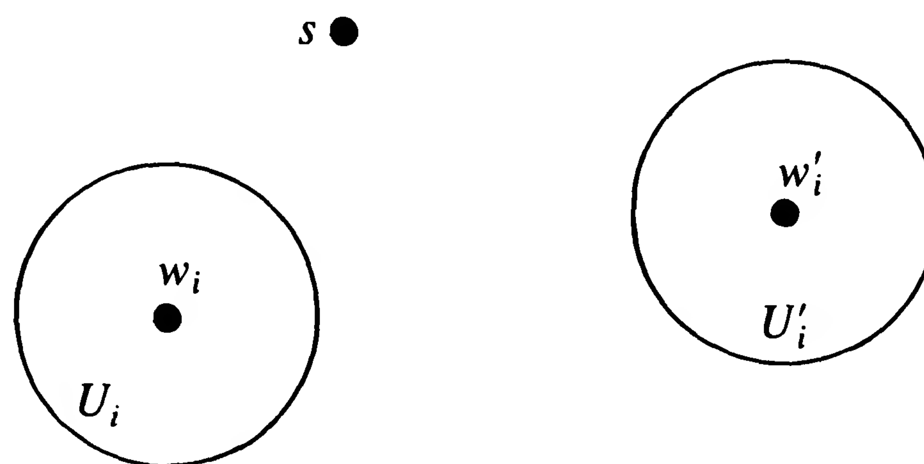
$$W = {}^t(w, 1), \; W' = {}^t(w', 1).$$

(b) Assume that $|\lambda| < |\lambda'|$. Given $z \neq w$, show that

$$\lim_{k \to \infty} M^k(z) = w'.$$

[*Hint*: Let $S = (W, W')$ and consider $S^{-1}M^kS(z) = \alpha^k z$ where $\alpha = \lambda/\lambda'$.]

56. (Tits) Let $M_1, \dots, M_r \in GL_2(\mathbf{C})$ be a finite number of matrices. Let λ_i, λ'_i be the eigenvalues of M_i. Assume that each M_i has two distinct complex fixed points, and that $|\lambda_i| < |\lambda'_i|$. Also assume that the fixed points for $M_1, \dots, M_r$ are all distinct from each other. Prove that there exists a positive integer k such that $M_1^k, \dots, M_r^k$ are the free generators of a free subgroup of $GL_2(\mathbf{C})$. [*Hint*: Let w_i, w'_i be the fixed points of M_i. Let U_i be a small disc centered at w_i and U'_i a small disc centered at w'_i. Let $S_i = U_i \cup U'_i$. Let s be a complex number which does not lie in any S_i. Let $G_i = \langle M_i^k \rangle$. Show that the conditions of Exercise 54 are satisfied for k sufficiently large.].



57. Let G be a group acting on a set X. Let Y be a subset of X. Let G_Y be the subset of G consisting of those elements g such that $gY \cap Y$ is not empty. Let $\overline{G}_Y$ be the subgroup of G generated by G_Y. Then $\overline{G}_Y Y$ and $(G - \overline{G}_Y)Y$ are disjoint. [*Hint*: Suppose that there exist $g_1 \in \overline{G}_Y$ and $g_2 \in G$ but $g_2 \notin \overline{G}_Y$, and elements $y_1, y_2, \in Y$ such that $g_2y_1 = g_2y_2$. Then $g_2^{-1}g_1y_1 = y_2$, so $g_2^{-1}g_1 \in G_Y$ whence $g_2 \in \overline{G}_Y$, contrary to assumption.]

Application. Suppose that $X = GY$, but that X cannot be expressed as a disjoint union as above unless one of the two sets is empty. Then we conclude that $G - \overline{G}_Y$ is empty, and therefore G_Y generates G.

Example 1. Suppose X is a connected topological space, Y is open, and G acts continuously. Then all translates of Y are open, so G is generated by G_Y.

Example 2. Suppose G is a discrete group acting continuously and discretely on X. Again suppose X connected and Y closed. Then any union of translates of Y by elements of G is closed, so again $G - \overline{G}_Y$ is empty, and G_Y generates G.

CHAPTER II

Rings

§1. RINGS AND HOMOMORPHISMS

A **ring** A is a set, together with two laws of composition called multiplication and addition respectively, and written as a product and as a sum respectively, satisfying the following conditions:

RI 1. With respect to addition, A is a commutative group.

RI 2. The multiplication is associative, and has a unit element.

RI 3. For all $x, y, z \in A$ we have

$$(x + y)z = xz + yz \qquad \text{and} \qquad z(x + y) = zx + zy.$$

(This is called **distributivity**.)

As usual, we denote the unit element for addition by 0, and the unit element for multiplication by 1. We do not assume that $1 \neq 0$. We observe that $0x = 0$ for all $x \in A$. *Proof*: We have $0x + x = (0 + 1)x = 1x = x$. Hence $0x = 0$. In particular, if $1 = 0$, then A consists of 0 alone.

For any $x, y \in A$ we have $(-x)y = -(xy)$. *Proof*: We have

$$xy + (-x)y = (x + (-x))y = 0y = 0,$$

so $(-x)y$ is the additive inverse of xy.

Other standard laws relating addition and multiplication are easily proved, for instance $(-x)(-y) = xy$. We leave these as exercises.

Let A be a ring, and let U be the set of elements of A which have both a right and left inverse. Then U is a multiplicative group. Indeed, if a has a

right inverse b, so that $ab = 1$, and a left inverse c, so that $ca = 1$, then $cab = b$, whence $c = b$, and we see that c (or b) is a two-sided inverse, and that c itself has a two-sided inverse, namely a. Therefore U satisfies all the axioms of a multiplicative group, and is called the group of **units** of A. It is sometimes denoted by A^*, and is also called the group of **invertible** elements of A. A ring A such that $1 \neq 0$, and such that every non-zero element is invertible is called a **division ring**.

Note. The elements of a ring which are *left* invertible do not necessarily form a group.

Example. (The Shift Operator). Let E be the set of all sequences

$$a = (a_1, a_2, a_3, \ldots)$$

of integers. One can define addition componentwise. Let R be the set of all mappings $f: E \to E$ of E into itself such that $f(a + b) = f(a) + f(b)$. The law of composition is defined to be composition of mappings. Then R is a ring. (Proof?) Let

$$T(a_1, a_2, a_3, \ldots) = (0, a_1, a_2, a_3, \ldots).$$

Verify that T is left invertible but not right invertible.

A ring A is said to be **commutative** if $xy = yx$ for all $x, y \in A$. A commutative division ring is called a **field**. We observe that by definition, a field contains at least two elements, namely 0 and 1.

A subset B of a ring A is called a **subring** if it is an additive subgroup, if it contains the multiplicative unit, and if $x, y \in B$ implies $xy \in B$. If that is the case, then B itself is a ring, the laws of operation in B being the same as the laws of operation in A.

For example, the **center** of a ring A is the subset of A consisting of all elements $a \in A$ such that $ax = xa$ for all $x \in A$. One sees immediately that the center of A is a subring.

Just as we proved general associativity from the associativity for three factors, one can prove general distributivity. If $x, y_1, \ldots, y_n$ are elements of a ring A, then by induction one sees that

$$x(y_1 + \cdots + y_n) = xy_1 + \cdots + xy_n.$$

If x_i $(i = 1, \ldots, n)$ and y_j $(j = 1, \ldots, m)$ are elements of A, then it is also easily proved that

$$\left(\sum_{i=1}^{n} x_i\right)\left(\sum_{j=1}^{m} y_j\right) = \sum_{i=1}^{n} \sum_{j=1}^{m} x_i x_j.$$

Furthermore, distributivity holds for subtraction, e.g.

$$x(y_1 - y_2) = xy_1 - xy_2.$$

We leave all the proofs to the reader.

Examples. *Let S be a set and A a ring. Let* Map(S, A) *be the set of mappings of S into A. Then* Map(S, A) *is a ring if for $f, g \in$* Map(S, A) *we define*

$$(fg)(x) = f(x)g(x) \qquad \textit{and} \qquad (f + g)(x) = f(x) + g(x)$$

for all $x \in S$. The multiplicative unit is the constant map whose value is the multiplicative unit of A. The additive unit is the constant map whose value is the additive unit of A, namely 0. The verification that Map(S, A) is a ring under the above laws of composition is trivial and left to the reader.

Let M be an additive abelian group, and let A be the set End(M) of group-homomorphisms of M into itself. We define addition in A to be the addition of mappings, and we define multiplication to be **composition** of mappings. Then it is trivially verified that A is a ring. Its unit element is of course the identity mapping. In general, A is not commutative.

Readers have no doubt met polynomials over a field previously. These provide a basic example of a ring, and will be defined officially for this book in §3.

Let K be a field. The set of $n \times n$ matrices with components in K is a ring. Its units consist of those matrices which are invertible, or equivalently have a non-zero determinant.

Let S be a set and R the set of real-valued functions on S. Then R is a commutative ring. Its units consist of those functions which are nowhere 0. This is a special case of the ring Map(S, A) considered above.

The convolution product. We shall now give examples of rings whose product is given by what is called convolution. Let G be a group and let K be a field. Denote by $K[G]$ the set of all formal linear combinations $\alpha = \sum a_x x$ with $x \in G$ and $a_x \in K$, such that all but a finite number of a_x are equal to 0. (See §3, and also Chapter III, §4.) If $\beta = \sum b_x x \in K[G]$, then one can define the product

$$\alpha\beta = \sum_{x \in G} \sum_{y \in G} a_x b_y xy = \sum_{z \in G} \left(\sum_{xy=z} a_x b_y \right) z.$$

With this product, the **group ring** $K[G]$ is a ring, which will be studied extensively in Chapter XVIII when G is a finite group. Note that $K[G]$ is commutative if and only if G is commutative. The second sum on the right above defines what is called a **convolution product**. If f, g are two functions on a group G, we define their **convolution** $f * g$ by

$$(f * g)(z) = \sum_{xy=z} f(x)g(y).$$

Of course this must make sense. If G is infinite, one may restrict this definition to functions which are 0 except at a finite number of elements. Exercise 12 will give an example (actually on a monoid) when another type of restriction allows for a finite sum on the right.

Example from analysis. In analysis one considers a situation as follows. Let $L^1 = L^1(\mathbf{R})$ be the space of functions which are absolutely integrable.

Given functions $f, g \in L^1$, one defines their **convolution product** $f * g$ by

$$(f * g)(x) = \int_{\mathbf{R}} f(x - y)g(y)\, dy.$$

Then this product satisfies all the axioms of a ring, except that there is no unit element. In the case of the group ring or the convolution of Exercise 12, there is a unit element. (What is it?) Note that the convolution product in the case of $L^1(\mathbf{R})$ is commutative, basically because $\mathbf{R}$ is a commutative additive group. More generally, let G be a locally compact group with a Haar measure μ. Then the convolution product is defined by the similar formula

$$(f * g)(x) = \int_G f(xy^{-1})g(y)\, d\mu(y).$$

After these examples, we return to the general theory of rings.

A **left ideal** $\mathfrak{a}$ in a ring A is a subset of A which is a subgroup of the additive group of A, such that $A\mathfrak{a} \subset \mathfrak{a}$ (and hence $A\mathfrak{a} = \mathfrak{a}$ since A contains 1). To define a right ideal, we require $\mathfrak{a}A = \mathfrak{a}$, and a **two-sided ideal** is a subset which is both a left and a right ideal. A two-sided ideal is called simply an **ideal** in this section. Note that (0) and A itself are ideals.

If A is a ring and $a \in A$, then Aa is a left ideal, called **principal**. We say that a is a generator of $\mathfrak{a}$ (over A). Similarly, AaA is a principal two-sided ideal if we define AaA to be the set of all sums $\sum x_i a y_i$ with $x_i, y_i \in A$. Cf. below the definition of the product of ideals. More generally, let $a_1, \ldots, a_n$ be elements of A. We denote by $(a_1, \ldots, a_n)$ the set of elements of A which can be written in the form

$$x_1 a_1 + \cdots + x_n a_n \qquad \text{with} \quad x_i \in A.$$

Then this set of elements is immediately verified to be a left ideal, and $a_1, \ldots, a_n$ are called **generators** of the left ideal.

If $\{\mathfrak{a}_i\}_{i \in I}$ is a family of ideals, then their intersection

$$\bigcap_{i \in I} \mathfrak{a}_i$$

is also an ideal. Similarly for left ideals. Readers will easily verify that if $\mathfrak{a} = (a_1, \ldots, a_n)$, then $\mathfrak{a}$ is the intersection of all left ideals containing the elements $a_1, \ldots, a_n$.

A ring A is said to be **commutative** if $xy = yx$ for all $x, y \in A$. In that case, every left or right ideal is two-sided.

A **commutative** ring such that every ideal is principal and such that $1 \neq 0$ is called a **principal** ring.

Examples. The integers $\mathbf{Z}$ form a ring, which is commutative. Let $\mathfrak{a}$ be an ideal $\neq \mathbf{Z}$ and $\neq 0$. If $n \in \mathfrak{a}$, then $-n \in \mathfrak{a}$. Let d be the smallest integer > 0 lying in $\mathfrak{a}$. If $n \in \mathfrak{a}$ then there exist integers q, r with $0 \leqq r < d$ such that

$$n = dq + r.$$

Since $\mathfrak{a}$ is an ideal, it follows that r lies in $\mathfrak{a}$, hence $r = 0$. Hence $\mathfrak{a}$ consists of all multiples qd of d, with $q \in \mathbf{Z}$, and $\mathbf{Z}$ *is a principal ring.*

A similar example is the ring of polynomials in one variable over a field, as will be proved in Chapter IV, also using the Euclidean algorithm.

Let R be the ring of algebraic integers in a number field K. (For definitions, see Chapter VII.) Then R is not necessarily principal, but let $\mathfrak{p}$ be a prime ideal, and let $R_{\mathfrak{p}}$ be the ring of all elements a/b with $a, b \in R$ and $b \notin \mathfrak{p}$. Then in algebraic number theory, it is shown that $R_{\mathfrak{p}}$ is principal, with one prime ideal $\mathfrak{m}_{\mathfrak{p}}$ consisting of all elements a/b as above but with $a \in \mathfrak{p}$. See Exercises 15, 16, and 17.

An example from analysis. Let A be the set of entire functions on the complex plane. Then A is a commutative ring, and every finitely generated ideal is principal. Given a discrete set of complex numbers $\{z_i\}$ and integers $m_i \geqq 0$, there exists an entire function f having zeros at z_i of multiplicity m_i and no other zeros. Every principal ideal is of the form Af for some such f. The group of units A^* in A consists of the functions which have no zeros. It is a nice exercise in analysis to prove the above statements (using the Weierstrass factorization theorem).

We now return to general notions. Let $\mathfrak{a}$, $\mathfrak{b}$ be ideals of A. We define $\mathfrak{a}\mathfrak{b}$ to be the set of all sums

$$x_1 y_1 + \cdots + x_n y_n$$

with $x_i \in \mathfrak{a}$ and $y_i \in \mathfrak{b}$. Then one verifies immediately that $\mathfrak{a}\mathfrak{b}$ is an ideal, and that the set of ideals forms a multiplicative monoid, the unit element being the ring itself. This unit element is called the **unit ideal**, and is often written (1). If $\mathfrak{a}$, $\mathfrak{b}$ are left ideals, we define their product $\mathfrak{a}\mathfrak{b}$ as above. It is also a left ideal, and if $\mathfrak{a}$, $\mathfrak{b}$, $\mathfrak{c}$ are left ideals, then we again have associativity: $(\mathfrak{a}\mathfrak{b})\mathfrak{c} = \mathfrak{a}(\mathfrak{b}\mathfrak{c})$.

If $\mathfrak{a}$, $\mathfrak{b}$ are left ideals of A, then $\mathfrak{a} + \mathfrak{b}$ (the sum being taken as additive subgroup of A) is obviously a left ideal. Similarly for right and two-sided ideals. Thus ideals also form a monoid under addition. We also have distributivity: If $\mathfrak{a}_1, \ldots, \mathfrak{a}_n$, $\mathfrak{b}$ are ideals of A, then clearly

$$\mathfrak{b}(\mathfrak{a}_1 + \cdots + \mathfrak{a}_n) = \mathfrak{b}\mathfrak{a}_1 + \cdots + \mathfrak{b}\mathfrak{a}_n,$$

and similarly on the other side. (However, the set of ideals does not form a ring!)

Let $\mathfrak{a}$ be a left ideal. Define $\mathfrak{a}A$ to be the set of all sums $a_1 x_1 + \cdots + a_n x_n$ with $a_i \in \mathfrak{a}$ and $x_i \in A$. Then $\mathfrak{a}A$ is an ideal (two-sided).

Suppose that A is commutative. Let $\mathfrak{a}$, $\mathfrak{b}$ be ideals. Then trivially

$$\mathfrak{a}\mathfrak{b} \subset \mathfrak{a} \cap \mathfrak{b},$$

but equality does not necessarily hold. However, as an exercise, prove that if $\mathfrak{a} + \mathfrak{b} = A$ then $\mathfrak{a}\mathfrak{b} = \mathfrak{a} \cap \mathfrak{b}$.

As should be known to the reader, the integers $\mathbf{Z}$ satisfy another property besides every ideal being principal, namely unique factorization into primes.

We shall discuss the general phenomenon in §4. Be it noted here only that if a ring A has the property of unique factorization into prime elements, and p is a prime element, then the ideal (p) is prime, and the ring $R_{(p)}$ (defined as above) is principal. See Exercise 6. Thus principal rings may be obtained in a natural way from rings which are not principal.

As Dedekind found out, some form of unique factorization can be recovered in some cases, replacing unique factorization into prime elements by unique factorization of (non-zero) ideals into prime ideals.

Example. There are cases when the non-zero ideals give rise to a group. Let $\mathfrak{o}$ be a subring of a field K such that every element of K is a quotient of elements of $\mathfrak{o}$; that is, of the form a/b with $a, b \in \mathfrak{o}$ and $b \neq 0$. By a **fractional ideal** $\mathfrak{a}$ we mean a non-zero additive subgroup of K such that $\mathfrak{o}\mathfrak{a} \subset \mathfrak{a}$ (and therefore $\mathfrak{o}\mathfrak{a} = \mathfrak{a}$ since $\mathfrak{o}$ contains the unit element); and such that there exists an element $c \in \mathfrak{o}$, $c \neq 0$, such that $c\mathfrak{a} \subset \mathfrak{o}$. We might say that a fractional ideal has bounded denominator. A **Dedekind ring** is a ring $\mathfrak{o}$ as above such that the fractional ideals form a group under multiplication. As proved in books on algebraic number theory, the ring of algebraic integers in a number field is a Dedekind ring. Do Exercise 14 to get the property of unique factorization into prime ideals. See Exercise 7 of Chapter VII for a sketch of this proof.

If $a \in K$, $a \neq 0$, then $\mathfrak{o}a$ is a fractional ideal, and such ideals are called **principal**. The principal fractional ideals form a subgroup. The factor group is called the **ideal class group**, or **Picard group** of $\mathfrak{o}$, and is denoted by $\text{Pic}(\mathfrak{o})$. See Exercises 13–19 for some elementary facts about Dedekind rings. It is a basic problem to determine $\text{Pic}(\mathfrak{o})$ for various Dedekind rings arising in algebraic number theory and function theory. See my book *Algebraic Number Theory* for the beginnings of the theory in number fields. In the case of function theory, one is led to questions in algebraic geometry, notably the study of groups of divisor classes on algebraic varieties and all that this entails. The property that the fractional ideals form a group is essentially associated with the ring having "dimension 1" (which we do not define here). In general one is led into the study of modules under various equivalence relations; see for instance the comments at the end of Chapter III, §4.

We return to the general theory of rings.

By a **ring-homomorphism** one means a mapping $f: A \to B$ where A, B are rings, and such that f is a monoid-homomorphism for the multiplicative structures on A and B, and also a monoid-homomorphism for the additive structure. In other words, f must satisfy:

$$f(a + a') = f(a) + f(a'), \qquad f(aa') = f(a)f(a'),$$
$$f(1) = 1, \qquad f(0) = 0,$$

for all $a, a' \in A$. Its **kernel** is defined to be the kernel of f viewed as additive homomorphism.

The kernel of a ring-homomorphism $f: A \to B$ is an ideal of A, as one verifies at once.

Conversely, let $\mathfrak{a}$ be an ideal of the ring A. We can construct the **factor ring** $A/\mathfrak{a}$ as follows. Viewing A and $\mathfrak{a}$ as additive groups, let $A/\mathfrak{a}$ be the factor group. We define a multiplicative law of composition on $A/\mathfrak{a}$: If $x + \mathfrak{a}$ and $y + \mathfrak{a}$ are two cosets of $\mathfrak{a}$, we define $(x + \mathfrak{a})(y + \mathfrak{a})$ to be the coset $(xy + \mathfrak{a})$. This coset is well defined, for if x_1, y_1 are in the same coset as x, y respectively, then one verifies at once that $x_1 y_1$ is in the same coset as xy. Our multiplicative law of composition is then obviously associative, has a unit element, namely the coset $1 + \mathfrak{a}$, and the distributive law is satisfied since it is satisfied for coset representatives. We have therefore defined a ring structure on $A/\mathfrak{a}$, and the canonical map

$$f: A \to A/\mathfrak{a}$$

is then clearly a ring-homomorphism.

If $g: A \to A'$ is a ring-homomorphism whose kernel contains $\mathfrak{a}$, then there exists a unique ring-homomorphism $g_*: A/\mathfrak{a} \to A'$ making the following diagram commutative:

$$\begin{array}{ccc} A & \xrightarrow{g} & A' \\ & {\scriptstyle f}\searrow \quad \nearrow{\scriptstyle g_*} & \\ & A/\mathfrak{a} & \end{array}$$

Indeed, viewing f, g as group-homomorphisms (for the additive structures), there is a unique group-homomorphism g_* making our diagram commutative. We contend that g_* is in fact a ring-homomorphism. We could leave the trivial proof to the reader, but we carry it out in full. If $x \in A$, then $g(x) = g_* f(x)$. Hence for $x, y \in A$,

$$\begin{aligned} g_*(f(x)f(y)) &= g_*(f(xy)) = g(xy) = g(x)g(y) \\ &= g_* f(x) g_* f(y). \end{aligned}$$

Given $\xi, \eta \in A/\mathfrak{a}$, there exist $x, y \in A$ such that $\xi = f(x)$ and $\eta = f(y)$. Since $f(1) = 1$, we get $g_* f(1) = g(1) = 1$, and hence the two conditions that g_* be a multiplicative monoid-homomorphism are satisfied, as was to be shown.

The statement we have just proved is equivalent to saying that the canonical map $f: A \to A/\mathfrak{a}$ is universal in the category of homomorphisms whose kernel contains $\mathfrak{a}$.

Let A be a ring, and denote its unit element by e for the moment. The map

$$\lambda: \mathbf{Z} \to A$$

such that $\lambda(n) = ne$ is a ring-homomorphism (obvious), and its kernel is an ideal (n), generated by an integer $n \geqq 0$. We have a canonical injective homomorphism $\mathbf{Z}/n\mathbf{Z} \to A$, which is a (ring) isomorphism between $\mathbf{Z}/n\mathbf{Z}$ and a

subring of A. If $n\mathbf{Z}$ is a prime ideal, then $n = 0$ or $n = p$ for some prime number p. In the first case, A contains as a subring a ring which is isomorphic to $\mathbf{Z}$, and which is often identified with $\mathbf{Z}$. In that case, we say that A has **characteristic** 0. If on the other hand $n = p$, then we say that A has **characteristic** p, and A contains (an isomorphic image of) $\mathbf{Z}/p\mathbf{Z}$ as a subring. We abbreviate $\mathbf{Z}/p\mathbf{Z}$ by $\mathbf{F}_p$.

If K is a field, then K has characteristic 0 or $p > 0$. In the first case, K contains as a subfield an isomorphic image of the rational numbers, and in the second case, it contains an isomorphic image of $\mathbf{F}_p$. In either case, this subfield will be called the **prime field** (contained in K). Since this prime field is the smallest subfield of K containing 1 and has no automorphism except the identity, it is customary to identify it with $\mathbf{Q}$ or $\mathbf{F}_p$ as the case may be. By the **prime ring** (in K) we shall mean either the integers $\mathbf{Z}$ if K has characteristic 0, or $\mathbf{F}_p$ if K has characteristic p.

Let A be a subring of a ring B. Let S be a subset of B commuting with A; in other words we have $as = sa$ for all $a \in A$ and $s \in S$. We denote by $A[S]$ the set of all elements

$$\sum a_{i_1 \cdots i_n} s_1^{i_1} \cdots s_n^{i_n},$$

the sum ranging over a finite number of n-tuples $(i_1, \ldots, i_n)$ of integers $\geqq 0$, and $a_{i_1 \cdots i_n} \in A$, $s_1, \ldots, s_n \in S$. If $B = A[S]$, we say that S is a set of **generators** (or more precisely, **ring generators**) for B over A, or that B is **generated** by S over A. If S is finite, we say that B is **finitely generated as a ring over** A. One might say that $A[S]$ consists of all not-necessarily-commutative polynomials in elements of S with coefficients in A. Note that elements of S may not commute with each other.

Example. The ring of matrices over a field is finitely generated over that field, but matrices don't necessarily commute.

As with groups, we observe that a homomorphism is uniquely determined by its effect on generators. In other words, let $f: A \to A'$ be a ring-homomorphism, and let $B = A[S]$ as above. Then there exists at most one extension of f to a ring-homomorphism of B having prescribed values on S.

Let A be a ring, $\mathfrak{a}$ an ideal, and S a subset of A. We write

$$S \equiv 0 \pmod{\mathfrak{a}}$$

if $S \subset \mathfrak{a}$. If $x, y \in A$, we write

$$x \equiv y \pmod{\mathfrak{a}}$$

if $x - y \in \mathfrak{a}$. If $\mathfrak{a}$ is principal, equal to (a), then we also write

$$x \equiv y \pmod{a}.$$

If $f: A \to A/\mathfrak{a}$ is the canonical homomorphism, then $x \equiv y \pmod{\mathfrak{a}}$ means that $f(x) = f(y)$. The congruence notation is sometimes convenient when we want to avoid writing explicitly the canonical map f.

The factor ring $A/\mathfrak{a}$ is also called a **residue class ring**. Cosets of $\mathfrak{a}$ in A are called **residue classes** modulo $\mathfrak{a}$, and if $x \in A$, then the coset $x + \mathfrak{a}$ is called the **residue class of x modulo** $\mathfrak{a}$.

We have defined the notion of an isomorphism in any category, and so a ring-isomorphism is a ring-homomorphism which has a two-sided inverse. As usual we have the criterion:

A ring-homomorphism $f: A \to B$ which is bijective is an isomorphism.

Indeed, there exists a set-theoretic inverse $g: B \to A$, and it is trivial to verify that g is a ring-homomorphism.

Instead of saying "ring-homomorphism" we sometimes say simply "homomorphism" if the reference to rings is clear. We note that rings form a category (the morphisms being the homomorphisms).

Let $f: A \to B$ be a ring-homomorphism. Then the image $f(A)$ of f is a subring of B. Proof obvious.

It is clear that an injective ring-homomorphism $f: A \to B$ establishes a ring-isomorphism between A and its image. Such a homomorphism will be called an **embedding** (of rings).

Let $f: A \to A'$ be a ring-homomorphism, and let $\mathfrak{a}'$ be an ideal of A'. Then $f^{-1}(\mathfrak{a}')$ is an ideal $\mathfrak{a}$ in A, and we have an induced injective homomorphism

$$A/\mathfrak{a} \to A'/\mathfrak{a}'.$$

The trivial proof is left to the reader.

Proposition 1.1. *Products exist in the category of rings.*

In fact, let $\{A_i\}_{i \in I}$ be a family of rings, and let $A = \prod A_i$ be their product as additive abelian groups. We define a multiplication in A in the obvious way: If $(x_i)_{i \in I}$ and $(y_i)_{i \in I}$ are two elements of A, we define their product to be $(x_i y_i)_{i \in I}$, i.e. we define multiplication componentwise, just as we did for addition. The multiplicative unit is simply the element of the product whose i-th component is the unit element of A_i. It is then clear that we obtain a ring structure on A, and that the projection on the i-th factor is a ring-homomorphism. Furthermore, A together with these projections clearly satisfies the required universal property.

Note that the usual inclusion of A_i on the i-th factor is *not* a ring-homomorphism because it does not map the unit element e_i of A_i on the unit element of A. Indeed, it maps e_i on the element of A having e_i as i-th component, and 0 $(= 0_i)$ as all other components.

Let A be a ring. Elements x, y of A are said to be **zero divisors** if $x \neq 0$, $y \neq 0$, and $xy = 0$. Most of the rings without zero divisors which we consider will be commutative. In view of this, we define a ring A to be **entire** if $1 \neq 0$, if A is commutative, and if there are no zero divisors in the ring. (Entire rings are also called **integral domains**. However, linguistically, I feel

the need for an adjective. "Integral" would do, except that in English, "integral" has been used for "integral over a ring" as in Chapter VII, §1. In French, as in English, two words exist with similar roots: "integral" and "entire". The French have used both words. Why not do the same in English? There is a slight psychological impediment, in that it would have been better if the use of "integral" and "entire" were reversed to fit the long-standing French use. I don't know what to do about this.)

Examples. The ring of integers $\mathbf{Z}$ is without zero divisors, and is therefore entire. If S is a set with more than 2 elements, and A is a ring with $1 \neq 0$, then the ring of mappings $\text{Map}(S, A)$ has zero divisors. (Proof?)

Let m be a positive integer $\neq 1$. The ring $\mathbf{Z}/m\mathbf{Z}$ has zero divisors if and only if m is not a prime number. (Proof left as an exercise.) The ring of $n \times n$ matrices over a field has zero divisors if $n \geqq 2$. (Proof?)

The next criterion is used very frequently.

Let A be an entire ring, and let a, b be non-zero elements of A. Then a, b generate the same ideal if and only if there exists a unit u of A such that $b = au$.

Proof. If such a unit exists we have $Ab = Aua = Aa$. Conversely, assume $Aa = Ab$. Then we can write $a = bc$ and $b = ad$ with some elements $c, d \in A$. Hence $a = adc$, whence $a(1 - dc) = 0$, and therefore $dc = 1$. Hence c is a unit.

§2. COMMUTATIVE RINGS

Throughout this section, we let A denote a commutative ring.

A **prime** ideal in A is an ideal $\mathfrak{p} \neq A$ such that $A/\mathfrak{p}$ is entire. Equivalently, we could say that it is an ideal $\mathfrak{p} \neq A$ such that, whenever $x, y \in A$ and $xy \in \mathfrak{p}$, then $x \in \mathfrak{p}$ or $y \in \mathfrak{p}$. A prime ideal is often called simply a **prime**.

Let $\mathfrak{m}$ be an ideal. We say that $\mathfrak{m}$ is a **maximal** ideal if $\mathfrak{m} \neq A$ and if there is no ideal $\mathfrak{a} \neq A$ containing $\mathfrak{m}$ and $\neq \mathfrak{m}$.

Every maximal ideal is prime.

Proof. Let $\mathfrak{m}$ be maximal and let $x, y \in A$ be such that $xy \in \mathfrak{m}$. Suppose $x \notin \mathfrak{m}$. Then $\mathfrak{m} + Ax$ is an ideal properly containing $\mathfrak{m}$, hence equal to A. Hence we can write

$$1 = u + ax$$

with $u \in \mathfrak{m}$ and $a \in A$. Multiplying by y we find

$$y = yu + axy,$$

whence $y \in \mathfrak{m}$ and $\mathfrak{m}$ is therefore prime.

Let $\mathfrak{a}$ be an ideal $\neq A$. Then $\mathfrak{a}$ is contained in some maximal ideal $\mathfrak{m}$.

Proof. The set of ideals containing $\mathfrak{a}$ and $\neq A$ is inductively ordered by ascending inclusion. Indeed, if $\{\mathfrak{b}_i\}$ is a totally ordered set of such ideals, then $1 \notin \mathfrak{b}_i$ for any i, and hence 1 does not lie in the ideal $\mathfrak{b} = \bigcup \mathfrak{b}_i$, which dominates all $\mathfrak{b}_i$. If $\mathfrak{m}$ is a maximal element in our set, then $\mathfrak{m} \neq A$ and $\mathfrak{m}$ is a maximal ideal, as desired.

The ideal $\{0\}$ is a prime ideal of A if and only if A is entire.

(Proof obvious.)

We defined a **field** K to be a commutative ring such that $1 \neq 0$, and such that the multiplicative monoid of non-zero elements of K is a group (i.e. such that whenever $x \in K$ and $x \neq 0$ then there exists an inverse for x). We note that the only ideals of a field K are K and the zero ideal.

If $\mathfrak{m}$ is a maximal ideal of A, then $A/\mathfrak{m}$ is a field.

Proof. If $x \in A$, we denote by $\bar{x}$ its residue class mod $\mathfrak{m}$. Since $\mathfrak{m} \neq A$ we note that $A/\mathfrak{m}$ has a unit element $\neq 0$. Any non-zero element of $A/\mathfrak{m}$ can be written as $\bar{x}$ for some $x \in A$, $x \notin \mathfrak{m}$. To find its inverse, note that $\mathfrak{m} + Ax$ is an ideal of $A \neq \mathfrak{m}$ and hence equal to A. Hence we can write

$$1 = u + yx$$

with $u \in \mathfrak{m}$ and $y \in A$. This means that $\bar{y}\bar{x} = 1$ (i.e. $= \bar{1}$) and hence that $\bar{x}$ has an inverse, as desired.

Conversely, we leave it as an exercise to the reader to prove that:

If $\mathfrak{m}$ is an ideal of A such that $A/\mathfrak{m}$ is a field, then $\mathfrak{m}$ is maximal.

Let $f: A \to A'$ be a homomorphism of commutative rings. Let $\mathfrak{p}'$ be a prime ideal of A', and let $\mathfrak{p} = f^{-1}(\mathfrak{p}')$. Then $\mathfrak{p}$ is prime.

To prove this, let $x, y \in A$, and $xy \in \mathfrak{p}$. Suppose $x \notin \mathfrak{p}$. Then $f(x) \notin \mathfrak{p}'$. But $f(x)f(y) = f(xy) \in \mathfrak{p}'$. Hence $f(y) \in \mathfrak{p}'$, as desired.

As an exercise, prove that if f is surjective, and if $\mathfrak{m}'$ is maximal in A', then $f^{-1}(\mathfrak{m}')$ is maximal in A.

Example. Let $\mathbf{Z}$ be the ring of integers. Since an ideal is also an additive subgroup of $\mathbf{Z}$, every ideal $\neq \{0\}$ is principal, of the form $n\mathbf{Z}$ for some integer $n > 0$ (uniquely determined by the ideal). Let $\mathfrak{p}$ be a prime ideal $\neq \{0\}$, $\mathfrak{p} = n\mathbf{Z}$. Then n must be a prime number, as follows essentially directly from the definition of a prime ideal. Conversely, if p is a prime number, then $p\mathbf{Z}$ is a prime ideal (trivial exercise). Furthermore, $p\mathbf{Z}$ is a maximal ideal. Indeed, suppose $p\mathbf{Z}$ contained in some ideal $n\mathbf{Z}$. Then $p = nm$ for some integer m, whence $n = p$ or $n = 1$, thereby proving $p\mathbf{Z}$ maximal.

If n is an integer, the factor ring $\mathbf{Z}/n\mathbf{Z}$ is called the ring of **integers modulo** n. We also denote

$$\mathbf{Z}/n\mathbf{Z} = \mathbf{Z}(n).$$

If n is a prime number p, then the ring of integers modulo p is in fact a field, denoted by $\mathbf{F}_p$. In particular, the multiplicative group of $\mathbf{F}_p$ is called the group of non-zero integers modulo p. From the elementary properties of groups, we get a standard fact of elementary number theory: If x is an integer $\not\equiv 0 \pmod{p}$, then $x^{p-1} \equiv 1 \pmod{p}$. (For simplicity, it is customary to write mod p instead of mod $p\mathbf{Z}$, and similarly to write mod n instead of mod $n\mathbf{Z}$ for any integer n.) Similarly, given an integer $n > 1$, the units in the ring $\mathbf{Z}/n\mathbf{Z}$ consist of those residue classes mod $n\mathbf{Z}$ which are represented by integers $m \neq 0$ and prime to n. The order of the group of units in $\mathbf{Z}/n\mathbf{Z}$ is called by definition $\varphi(n)$ (where φ is known as the **Euler phi-function**). Consequently, if x is an integer prime to n, then $x^{\varphi(n)} \equiv 1 \pmod{n}$.

Theorem 2.1. (Chinese Remainder Theorem). *Let $\mathfrak{a}_1, \ldots, \mathfrak{a}_n$ be ideals of A such that $\mathfrak{a}_i + \mathfrak{a}_j = A$ for all $i \neq j$. Given elements $x_1, \ldots, x_n \in A$, there exists $x \in A$ such that $x \equiv x_i \pmod{\mathfrak{a}_i}$ for all i.*

Proof. If $n = 2$, we have an expression

$$1 = a_1 + a_2$$

for some elements $a_i \in \mathfrak{a}_i$, and we let $x = x_2 a_1 + x_1 a_2$.
For each $i \geqq 2$ we can find elements $a_i \in \mathfrak{a}_1$ and $b_i \in \mathfrak{a}_i$ such that

$$a_i + b_i = 1, \qquad i \geqq 2.$$

The product $\prod_{i=2}^{n} (a_i + b_i)$ is equal to 1, and lies in

$$\mathfrak{a}_1 + \prod_{i=2}^{n} \mathfrak{a}_i,$$

i.e. in $\mathfrak{a}_1 + \mathfrak{a}_2 \cdots \mathfrak{a}_n$. Hence

$$\mathfrak{a}_1 + \prod_{i=2}^{n} \mathfrak{a}_i = A.$$

By the theorem for $n = 2$, we can find an element $y_1 \in A$ such that

$$y_1 \equiv 1 \quad (\text{mod } \mathfrak{a}_1),$$

$$y_1 \equiv 0 \quad \left(\text{mod } \prod_{i=2}^{n} \mathfrak{a}_i\right).$$

We find similarly elements $y_2, \ldots, y_n$ such that

$$y_j \equiv 1 \pmod{\mathfrak{a}_j} \qquad \text{and} \qquad y_j \equiv 0 \pmod{\mathfrak{a}_i} \qquad \text{for } i \neq j.$$

Then $x = x_1 y_1 + \cdots + x_n y_n$ satisfies our requirements.

In the same vein as above, we observe that if $\mathfrak{a}_1, \dots, \mathfrak{a}_n$ are ideals of a ring A such that

$$\mathfrak{a}_1 + \cdots + \mathfrak{a}_n = A,$$

and if $\nu_1, \dots, \nu_n$ are positive integers, then

$$\mathfrak{a}_1^{\nu_1} + \cdots + \mathfrak{a}_n^{\nu_n} = A.$$

The proof is trivial, and is left as an exercise.

Corollary 2.2. *Let $\mathfrak{a}_1, \dots, \mathfrak{a}_n$ be ideals of A. Assume that $\mathfrak{a}_i + \mathfrak{a}_j = A$ for $i \neq j$. Let*

$$f: A \to \prod_{i=1}^{n} A/\mathfrak{a}_i = (A/\mathfrak{a}_1) \times \cdots \times (A/\mathfrak{a}_n)$$

be the map of A into the product induced by the canonical map of A onto $A/\mathfrak{a}_i$ for each factor. Then the kernel of f is $\bigcap_{i=1}^{n} \mathfrak{a}_i$, and f is surjective, thus giving an isomorphism

$$A/\bigcap \mathfrak{a}_i \xrightarrow{\approx} \prod A/\mathfrak{a}_i.$$

Proof. That the kernel of f is what we said it is, is obvious. The surjectivity follows from the theorem.

The theorem and its corollary are frequently applied to the ring of integers $\mathbf{Z}$ and to distinct prime ideals $(p_1), \dots, (p_n)$. These satisfy the hypothesis of the theorem since they are maximal. Similarly, one could take integers $m_1, \dots, m_n$ which are relatively prime in pairs, and apply the theorem to the principal ideals $(m_1) = m_1\mathbf{Z}, \dots, (m_n) = m_n\mathbf{Z}$. This is the ultraclassical case of the Chinese remainder theorem.

In particular, let m be an integer > 1, and let

$$m = \prod_i p_i^{r_i}$$

be a factorization of m into primes, with exponents $r_i \geqq 1$. Then we have a ring-isomorphism:

$$\mathbf{Z}/m\mathbf{Z} \approx \prod_i \mathbf{Z}/p_i^{r_i}\mathbf{Z}.$$

If A is a ring, we denote as usual by A^* the multiplicative group of invertible elements of A. We leave the following assertions as exercises:

The preceding ring-isomorphism of $\mathbf{Z}/m\mathbf{Z}$ onto the product induces a group-isomorphism

$$(\mathbf{Z}/m\mathbf{Z})^* \approx \prod_i (\mathbf{Z}/p_i^{r_i}\mathbf{Z})^*.$$

In view of our isomorphism, we have

$$\varphi(m) = \prod_i \varphi(p_i^{r_i}).$$

If p is a prime number and r an integer $\geqq 1$, then

$$\varphi(p^r) = (p-1)p^{r-1}.$$

One proves this last formula by induction. If $r = 1$, then $\mathbf{Z}/p\mathbf{Z}$ is a field, and the multiplicative group of that field has order $p - 1$. Let r be $\geqq 1$, and consider the canonical ring-homomorphism

$$\mathbf{Z}/p^{r+1}\mathbf{Z} \to \mathbf{Z}/p^r\mathbf{Z},$$

arising from the inclusion of ideals $(p^{r+1}) \subset (p^r)$. We get an induced group-homomorphism

$$\lambda: (\mathbf{Z}/p^{r+1}\mathbf{Z})^* \to (\mathbf{Z}/p^r\mathbf{Z})^*,$$

which is surjective because any integer a which represents an element of $\mathbf{Z}/p^r\mathbf{Z}$ and is prime to p will represent an element of $(\mathbf{Z}/p^{r+1}\mathbf{Z})^*$. Let a be an integer representing an element of $(\mathbf{Z}/p^{r+1}\mathbf{Z})^*$, such that $\lambda(a) = 1$. Then

$$a \equiv 1 \pmod{p^r\mathbf{Z}},$$

and hence we can write

$$a \equiv 1 + xp^r \pmod{p^{r+1}\mathbf{Z}}$$

for some $x \in \mathbf{Z}$. Letting $x = 0, 1, \ldots, p - 1$ gives rise to p distinct elements of $(\mathbf{Z}/p^{r+1}\mathbf{Z})^*$, all of which are in the kernel of λ. Furthermore, the element x above can be selected to be one of these p integers because every integer is congruent to one of these p integers modulo (p). Hence the kernel of λ has order p, and our formula is proved.

Note that the kernel of λ is isomorphic to $\mathbf{Z}/p\mathbf{Z}$. (Proof?)

Application: The ring of endomorphisms of a cyclic group. One of the first examples of a ring is the ring of endomorphisms of an abelian group. In the case of a cyclic group, we have the following complete determination.

Theorem 2.3. *Let A be a cyclic group of order n. For each $k \in \mathbf{Z}$ let $f_k: A \to A$ be the endomorphism $x \mapsto kx$ (writing A additively). Then $k \mapsto f_k$ induces a ring isomorphism $\mathbf{Z}/n\mathbf{Z} \approx \mathrm{End}(A)$, and a group isomorphism $(\mathbf{Z}/n\mathbf{Z})^* \approx \mathrm{Aut}(A)$.*

Proof. Recall that the additive group structure on $\mathrm{End}(A)$ is simply addition of mappings, and the multiplication is composition of mappings. The fact that $k \mapsto f_k$ is a ring-homomorphism is then a restatement of the formulas

$$1a = a, \qquad (k + k')a = ka + k'a, \qquad \text{and} \qquad (kk')a = k(k'a)$$

for $k, k' \in \mathbf{Z}$ and $a \in A$. If a is a generator of A, then $ka = 0$ if and only if $k \equiv 0 \bmod n$, so $\mathbf{Z}/n\mathbf{Z}$ is embedded in $\mathrm{End}(A)$. On the other hand, let $f: A \to A$ be an endomorphism. Again for a generator a, we have $f(a) = ka$

for some k, whence $f = f_k$ since every $x \in A$ is of the form ma for some $m \in Z$, and

$$f(x) = f(ma) = mf(a) = mka = kma = kx.$$

This proves the isomorphism $\mathbf{Z}/n\mathbf{Z} \approx \mathrm{End}(A)$. Furthermore, if $k \in (\mathbf{Z}/n\mathbf{Z})^*$ then there exists k' such that $kk' \equiv 1 \bmod n$, so f_k has the inverse $f_{k'}$ and f_k is an automorphism. Conversely, given any automorphism f with inverse g, we know from the first part of the proof that $f = f_k$, $g = g_{k'}$ for some k, k', and $f \circ g = \mathrm{id}$ means that $kk' \equiv 1 \bmod n$, so $k, k' \in (\mathbf{Z}/n\mathbf{Z})^*$. This proves the isomorphism $(\mathbf{Z}/n\mathbf{Z})^* \approx \mathrm{Aut}(A)$.

Note that if A is written as a multiplicative group C, then the map f_k is given by $x \mapsto x^k$. For instance, let $\boldsymbol{\mu}_n$ be the group of n-th roots of unity in $\mathbf{C}$. Then all automorphisms of $\boldsymbol{\mu}_n$ are given by

$$\zeta \mapsto \zeta^k \quad \text{with} \quad k \in (\mathbf{Z}/n\mathbf{Z})^*.$$

§3. POLYNOMIALS AND GROUP RINGS

Although all readers will have met polynomial functions, this section lays the ground work for polynomials in general. One needs polynomials over arbitrary rings in many contexts. For one thing, there are polynomials over a finite field which cannot be identified with polynomial functions in that field. One needs polynomials with integer coefficients, and one needs to reduce these polynomials mod p for primes p. One needs polynomials over arbitrary commutative rings, both in algebraic geometry and in analysis, for instance the ring of polynomial differential operators. We also have seen the example of a ring $B = A[S]$ generated by a set of elements over a ring A. We now give a systematic account of the basic definitions of polynomials over a commutative ring A.

We want to give a meaning to an expression such as

$$a_0 + a_1X + \cdots + a_nX^n,$$

where $a_i \in A$ and X is a "variable". There are several devices for doing so, and we pick one of them. (I picked another in my *Undergraduate Algebra.*) Consider an infinite cyclic group generated by an element X. We let S be the subset consisting of powers X^r with $r \geqq 0$. Then S is a monoid. We define the set of **polynomials** $A[X]$ to be the set of functions $S \to A$ which are equal to 0 except for a finite number of elements of S. For each element $a \in A$ we denote by aX^n the function which has the value a on X^n and the value 0 for all other elements of S. Then it is immediate that a polynomial can be written uniquely as a finite sum

$$a_0X^0 + \cdots + a_nX^n$$

for some integer $n \in \mathbf{N}$ and $a_i \in A$. Such a polynomial is denoted by $f(X)$. The elements $a_i \in A$ are called the **coefficients** of f. We define the product according to the convolution rule. Thus, given polynomials

$$f(X) = \sum_{i=0}^{n} a_iX^i \qquad \text{and} \qquad g(X) = \sum_{j=0}^{m} b_jX^j$$

we define the product to be

$$f(X)g(X) = \sum_{k=0}^{m+n} \left(\sum_{i+j=k} a_ib_j \right) X^k.$$

It is immediately verified that this product is associative and distributive. We shall give the details of associativity in the more general context of a monoid ring below. Observe that there is a unit element, namely $1X^0$. There is also an embedding

$$A \to A[X] \qquad \text{given by} \qquad a \mapsto aX^0.$$

One usually does not distinguish a from its image in $A[X]$, and one writes a instead of aX^0. Note that for $c \in A$ we have then $cf(x) = \sum ca_iX^i$.

Observe that by our definition, we have an equality of polynomials

$$\sum a_iX^i = \sum b_iX^i$$

if and only if $a_i = b_i$ for all i.

Let A be a subring of a commutative ring B. Let $x \in B$. If $f \in A[X]$ is a polynomial, we may then define the associated **polynomial function**

$$f_B: B \to B$$

by letting

$$f_B(x) = f(x) = a_0 + a_1x + \cdots + a_nx^n.$$

Given an element $b \in B$, directly from the definition of multiplication of polynomials, we find:

The association

$$\mathrm{ev}_b: f \mapsto f(b)$$

is a ring homomorphism of $A[X]$ into B.

This homomorphism is called the **evaluation homomorphism**, and is also said to be obtained by **substituting** b for X in the polynomial. (Cf. Proposition 3.1 below.)

Let $x \in B$. We now see that the subring $A[x]$ of B generated by x over A is the ring of all polynomial values $f(x)$, for $f \in A[X]$. If the evaluation map $f \mapsto f(x)$ gives an isomorphism of $A[X]$ with $A[x]$, then we say that x is

transcendental over A, or that x is a **variable** over A. In particular, X is a variable over A.

Example. Let $\alpha = \sqrt{2}$. Then the set of all real numbers of the form $a + b\alpha$, with $a, b \in \mathbf{Z}$, is a subring of the real numbers, generated by $\sqrt{2}$. Note that α is not transcendental over $\mathbf{Z}$, because the polynomial $X^2 - 2$ lies in the kernel of the evaluation map $f \mapsto f(\sqrt{2})$. On the other hand, it can be shown that $e = 2.718\ldots$ and π are transcendental over $\mathbf{Q}$. See Appendix 1.

Example. Let p be a prime number and let $K = \mathbf{Z}/p\mathbf{Z}$. Then K is a field. Let $f(X) = X^p - X \in K[X]$. Then f is not the zero polynomial. But f_K is the zero function. Indeed, $f_K(0) = 0$. If $x \in K$, $x \neq 0$, then since the multiplicative group of K has order $p - 1$, it follows that $x^{p-1} = 1$, whence $x^p = x$, so $f(x) = 0$. Thus a non-zero polynomial gives rise to the zero function on K.

There is another homomorphism of the polynomial ring having to do with the coefficients. Let

$$\varphi: A \to B$$

be a homomorphism of commutative rings. Then there is an associated homomorphism of the polynomial rings $A[X] \to B[X]$, such that

$$f(X) = \sum a_i X^i \mapsto \sum \varphi(a_i) X^i = (\varphi f)(X).$$

The verification that this mapping is a homomorphism is immediate, and further details will be given below in Proposition 3.2, in a more general context. We call $f \mapsto \varphi f$ the **reduction map**.

Examples. In some applications the map φ may be an isomorphism. For instance, if $f(X)$ has complex coefficients, then its complex conjugate $\bar{f}(X) = \sum \bar{a}_i X^i$ is obtained by applying complex conjugation to its coefficients.

Let $\mathfrak{p}$ be a prime ideal of A. Let $\varphi: A \to A'$ be the canonical homomorphism of A onto $A/\mathfrak{p}$. If $f(X)$ is a polynomial in $A[X]$, then φf will sometimes be called the **reduction of f modulo $\mathfrak{p}$**.

For example, taking $A = \mathbf{Z}$ and $\mathfrak{p} = (p)$ where p is a prime number, we can speak of the polynomial $3X^4 - X + 2$ as a polynomial mod 5, viewing the coefficients 3, -1, 2 as integers mod 5, i.e. elements of $\mathbf{Z}/5\mathbf{Z}$.

We may now combine the evaluation map and the reduction map to generalize the evaluation map.

Let $\varphi: A \to B$ be a homomorphism of commutative rings.
Let $x \in B$. There is a unique homomorphism extending φ

$$A[X] \to B \qquad \textit{such that} \qquad X \mapsto x,$$

and for this homomorphism, $\sum a_i X^i \mapsto \sum \varphi(a_i) x^i$.

The homomorphism of the above statement may be viewed as the composite

$$A[X] \longrightarrow B[X] \xrightarrow{\text{ev}_x} B$$

where the first map applies φ to the coefficients of a polynomial, and the second map is the evaluation at x previously discussed.

Example. In Chapter IX, §2 and §3, we shall discuss such a situation in several variables, when $(\varphi f)(x) = 0$, in which case x is called a **zero** of the polynomial f.

When writing a polynomial $f(X) = \sum_{i=1}^{n} a_i X^i$, if $a_n \neq 0$ then we define n to be the **degree** of f. Thus the degree of f is the smallest integer n such that $a_r = 0$ for $r > n$. If $f = 0$ (i.e. f is the zero polynomial), then by convention we define the degree of f to be $-\infty$. We agree to the convention that

$$-\infty + -\infty = -\infty, \qquad -\infty + n = -\infty, \qquad -\infty < n,$$

for all $n \in \mathbf{Z}$, and no other operation with $-\infty$ is defined. A polynomial of degree 1 is also called a **linear** polynomial. If $f \neq 0$ and $\deg f = n$, then we call a_n the **leading coefficient** of f. We call a_0 its **constant term.**

Let

$$g(X) = b_0 + \cdots + b_m X^m$$

be a polynomial in $A[X]$, of degree m, and assume $g \neq 0$. Then

$$f(X)g(X) = a_0 b_0 + \cdots + a_n b_m X^{m+n}.$$

Therefore:

If we assume that at least one of the leading coefficients a_n or b_m is not a divisor of 0 in A, then

$$\deg(fg) = \deg f + \deg g$$

and the leading coefficient of fg is $a_n b_m$. This holds in particular when a_n or b_m is a unit in A, or when A is entire. Consequently, when A is entire, $A[X]$ is also entire.

If f or $g = 0$, then we still have

$$\deg(fg) = \deg f + \deg g$$

if we agree that $-\infty + m = -\infty$ for any integer m.

One verifies trivially that for any polynomial $f, g \in A[X]$ we have

$$\deg(f + g) \leqq \max(\deg f, \deg g),$$

again agreeing that $-\infty < m$ for every integer m.

Polynomials in several variables

We now go to polynomials in several variables. Let A be a subring of a commutative ring B. Let $x_1, \dots, x_n \in B$. For each n-tuple of integers $(\nu_1, \dots, \nu_n) = (\nu) \in \mathbf{N}^n$, we use vector notation, letting $(x) = (x_1, \dots, x_n)$, and

$$M_{(\nu)}(x) = x_1^{\nu_1} \cdots x_n^{\nu_n}.$$

The set of such elements forms a monoid under multiplication. Let $A[x] = A[x_1, \dots, x_n]$ be the subring of B generated by $x_1, \dots, x_n$ over A. Then every element of $A[x]$ can be written as a finite sum

$$\sum a_{(\nu)} M_{(\nu)}(x) \qquad \text{with} \quad a_{(\nu)} \in A.$$

Using the construction of polynomials in one variable repeatedly, we may form the ring

$$A[X_1, \dots, X_n] = A[X_1][X_2] \cdots [X_n],$$

selecting X_n to be a variable over $A[X_1, \dots, X_{n-1}]$. Then every element f of $A[X_1, \dots, X_n] = A[X]$ has a *unique* expression as a finite sum

$$f = \sum_{j=0}^{d_n} f_j(X_1, \dots, X_{n-1}) X_n^j \qquad \text{with} \quad f_j \in A[X_1, \dots, X_{n-1}].$$

Therefore by induction we can write f uniquely as a sum

$$\begin{aligned} f &= \sum_{\nu_n = 0}^{d_n} \left(\sum_{\nu_1, \dots, \nu_{n-1}} a_{\nu_1 \cdots \nu_n} X_1^{\nu_1} \cdots X_{n-1}^{\nu_{n-1}} \right) X_n^{\nu_n} \\ &= \sum a_{(\nu)} M_{(\nu)}(X) = \sum a_{(\nu)} X_1^{\nu_1} \cdots X_n^{\nu_n} \end{aligned}$$

with elements $a_{(\nu)} \in A$, which are called the **coefficients** of f. The products

$$M_{(\nu)}(X) = X_1^{\nu_1} \cdots X_n^{\nu_n}$$

will be called **primitive monomials**. Elements of $A[X]$ are called **polynomials** (in n variables). We call $a_{(\nu)}$ its **coefficients**.

Just as in the one-variable case, we have an evaluation map. Given $(x) = (x_1, \dots, x_n)$ and f as above, we define

$$f(x) = \sum a_{(\nu)} M_{(\nu)}(x) = \sum a_{(\nu)} x_1^{\nu_1} \cdots x_n^{\nu_n}.$$

Then the **evaluation map**

$$\mathrm{ev}_{(x)} : A[X] \to B \qquad \text{such that} \qquad f \mapsto f(x)$$

is a ring-homomorphism. It may be viewed as the composite of the successive evaluation maps in one variable $X_i \mapsto x_i$ for $i = n, \dots, 1$, because $A[X] \subset B[X]$.

Just as for one variable, if $f(X) \in A[X]$ is a polynomial in n variables, then we obtain a function

$$f_B: B^n \to B \qquad \text{by} \qquad (x) \mapsto f(x).$$

We say that $f(x)$ is obtained by **substituting** (x) for (X) in f, or by **specializing** (X) to (x). As for one variable, if K is a finite field, and $f \in K[X]$ one may have $f \neq 0$ but $f_K = 0$. Cf. Chapter IV, Theorem 1.4 and its corollaries.

Next let $\varphi: A \to B$ be a homomorphism of commutative rings. Then we have the **reduction map** (generalized in Proposition 3.2 below)

$$f(X) = \sum a_{(\nu)} M_{(\nu)}(X) \mapsto \sum \varphi(a_{(\nu)}) M_{(\nu)}(X) = (\varphi f)(X).$$

We can also compose the evaluation and reduction. An element $(x) \in B^n$ is called a **zero** of f if $(\varphi f)(x) = 0$. Such zeros will be studied in Chapter IX.

Go back to A as a subring of B. Elements $x_1, \ldots, x_n \in B$ are called **algebraically independent** over A if the evaluation map

$$f \mapsto f(x)$$

is injective. Equivalently, we could say that if $f \in A[X]$ is a polynomial and $f(x) = 0$, then $f = 0$; in other words, there are no non-trivial polynomial relations among $x_1, \ldots, x_n$ over A.

Example. It is not known if e and π are algebraically independent over the rationals. It is not even known if $e + \pi$ is rational.

We now come to the notion of degree for several variables. By the **degree** of a primitive monomial

$$M_{(\nu)}(X) = X_1^{\nu_1} \cdots X_n^{\nu_n}$$

we shall mean the integer $|\nu| = \nu_1 + \cdots + \nu_n$ (which is $\geqq 0$).

A polynomial

$$aX_1^{\nu_1} \cdots X_n^{\nu_n} \qquad (a \in A)$$

will be called a **monomial** (not necessarily primitive).

If $f(X)$ is a polynomial in $A[X]$ written as

$$f(X) = \sum a_{(\nu)} X_1^{\nu_1} \cdots X_n^{\nu_n},$$

then either $f = 0$, in which case we say that its degree is $-\infty$, or $f \neq 0$, and then we define the **degree** of f to be the maximum of the degrees of the monomials $M_{(\nu)}(X)$ such that $a_{(\nu)} \neq 0$. (Such monomials are said to **occur** in the polynomial.) We note that the degree of f is 0 if and only if

$$f(X) = a_0 X_1^0 \cdots X_n^0$$

for some $a_0 \in A$, $a_0 \neq 0$. We also write this polynomial simply $f(X) = a_0$, i.e. writing 1 instead of

$$X_1^0 \cdots X_n^0,$$

in other words, we identify the polynomial with the constant a_0.

Note that a polynomial $f(X_1, \dots, X_n)$ in n variables can be viewed as a polynomial in X_n with coefficients in $A[X_1, \dots, X_{n-1}]$ (if $n \geqq 2$). Indeed, we can write

$$f(X) = \sum_{j=0}^{d_n} f_j(X_1, \dots, X_{n-1})X_n^j,$$

where f_j is an element of $A[X_1, \dots, X_{n-1}]$. By the **degree of f in X_n** we shall mean its degree when viewed as a polynomial in X_n with coefficients in $A[X_1, \dots, X_{n-1}]$. One sees easily that if this degree is d, then d is the largest integer occurring as an exponent of X_n in a monomial

$$a_{(\nu)} X_1^{\nu_1} \cdots X_n^{\nu_n}$$

with $a_{(\nu)} \neq 0$. Similarly, we define the degree of f in each variable X_i ($i = 1, \dots, n$).

The degree of f in each variable is of course usually different from its degree (which is sometimes called the **total degree** if there is need to prevent ambiguity). For instance,

$$X_1^3 X_2 + X_2^2$$

has total degree 4, and has degree 3 in X_1 and 2 in X_2.

As a matter of notation, we shall often abbreviate "degree" by "deg."

For each integer $d \geqq 0$, given a polynomial f, let $f^{(d)}$ be the sum of all monomials occurring in f and having degree d. Then

$$f = \sum_d f^{(d)}.$$

Suppose $f \neq 0$. We say that f is **homogeneous** of degree d if $f = f^{(d)}$; thus f can be written in the form

$$f(X) = \sum a_{(\nu)} X_1^{\nu_1} \cdots X_n^{\nu_n} \quad \text{with} \quad \nu_1 + \cdots + \nu_n = d \quad \text{if} \quad a_{(\nu)} \neq 0.$$

We shall leave it as an exercise to prove that *a non-zero polynomial f in n variables over A is homogeneous of degree d if and only if, for every set of $n + 1$ algebraically independent elements $u, t_1, \dots, t_n$ over A we have*

$$f(ut_1, \dots, ut_n) = u^d f(t_1, \dots, t_n).$$

We note that if f, g are homogeneous of degree d, e respectively, and $fg \neq 0$, then fg is homogeneous of degree $d + e$. If $d = e$ and $f + g \neq 0$, then $f + g$ is homogeneous of degree d.

Remark. In view of the isomorphism

$$A[X_1, \dots, X_n] \approx A[t_1, \dots, t_n]$$

between the polynomial ring in n variables and a ring generated over A by n

algebraically independent elements, we can apply all the terminology we have defined for polynomials, to elements of $A[t_1, \ldots, t_n]$. Thus we can speak of the degree of an element in $A[t]$, and the rules for the degree of a product or sum hold. In fact, we shall also call elements of $A[t]$ polynomials in (t). Algebraically independent elements will also be called **variables** (or independent variables), and any distinction which we make between $A[X]$ and $A[t]$ is more psychological than mathematical.

Suppose next that A is entire. By what we know of polynomials in one variable and induction, it follows that $A[X_1, \ldots, X_n]$ is entire. In particular, suppose f has degree d and g has degree e. Write

$$f = f^{(d)} + \text{terms of lower degree},$$

$$g = g^{(e)} + \text{terms of lower degree}.$$

Then $fg = f^{(d)}g^{(e)} + \text{terms of lower degree}$, and if $fg \neq 0$ then $f^{(d)}g^{(e)} \neq 0$. Thus we find:

$$\deg(fg) = \deg f + \deg g,$$

$$\deg(f + g) \leqq \max(\deg f, \deg g).$$

We are now finished with the basic terminology of polynomials. We end this section by indicating how the construction of polynomials is actually a special case of another construction which is used in other contexts. Interested readers can skip immediately to Chapter IV, giving further important properties of polynomials. See also Exercise 33 of Chapter XIII for harmonic polynomials.

The group ring or monoid ring

Let A be a commutative ring. Let G be a monoid, written multiplicatively.

Let $A[G]$ be the set of all maps $\alpha: G \to A$ such that $\alpha(x) = 0$ for almost all $x \in G$. We define addition in $A[G]$ to be the ordinary addition of mappings into an abelian (additive) group. If α, $\beta \in A[G]$, we define their product $\alpha\beta$ by the rule

$$(\alpha\beta)(z) = \sum_{xy=z} \alpha(x)\beta(y).$$

The sum is taken over all pairs (x, y) with x, $y \in G$ such that $xy = z$. This sum is actually finite, because there is only a finite number of pairs of elements $(x, y) \in G \times G$ such that $\alpha(x)\beta(y) \neq 0$. We also see that $(\alpha\beta)(t) = 0$ for almost all t, and thus belongs to our set $A[G]$.

The axioms for a ring are trivially verified. We shall carry out the proof of associativity as an example. Let α, β, $\gamma \in A[G]$. Then

$$\begin{aligned}((\alpha\beta)\gamma)(z) &= \sum_{xy=z} (\alpha\beta)(x)\gamma(y)\\ &= \sum_{xy=z}\left[\sum_{uv=x} \alpha(u)\beta(v)\right]\gamma(y)\\ &= \sum_{xy=z}\left[\sum_{uv=x} \alpha(u)\beta(v)\gamma(y)\right]\\ &= \sum_{\substack{(u,v,y)\\ uvy=z}} \alpha(u)\beta(v)\gamma(y),\end{aligned}$$

this last sum being taken over all triples $(u\ v,\ y)$ whose product is z. This last sum is now symmetric, and if we had computed $(\alpha(\beta\gamma))(z)$, we would have found this sum also. This proves associativity.

The unit element of $A[G]$ is the function δ such that $\delta(e) = 1$ and $\delta(x) = 0$ for all $x \in G$, $x \neq e$. It is trivial to verify that $\alpha = \delta\alpha = \alpha\delta$ for all $\alpha \in A[G]$.

We shall now adopt a notation which will make the structure of $A[G]$ clearer. Let $a \in A$ and $x \in G$. We denote by $a \cdot x$ (and sometimes also by ax) the function whose value at x is a, and whose value at y is 0 if $y \neq x$. Then an element $\alpha \in A[G]$ can be written as a sum

$$\alpha = \sum_{x \in G} \alpha(x) \cdot x.$$

Indeed, if $\{a_x\}_{x \in G}$ is a set of elements of A almost all of which are 0, and we set

$$\beta = \sum_{x \in G} a_x \cdot x,$$

then for any $y \in G$ we have $\beta(y) = a_y$ (directly from the definitions). This also shows that a given element α admits a unique expression as a sum $\sum a_x \cdot x$.

With our present notation, multiplication can be written

$$\left(\sum_{x \in G} a_x \cdot x\right)\left(\sum_{y \in G} b_y \cdot y\right) = \sum_{x,y} a_x b_y \cdot xy$$

and addition can be written

$$\sum_{x \in G} a_x \cdot x + \sum_{x \in G} b_x \cdot x = \sum_{x \in G} (a_x + b_x) \cdot x,$$

which looks the way we want it to look. Note that the unit element of $A[G]$ is simply $1 \cdot e$.

We shall now see that we can embed both A and G in a natural way in $A[G]$.

Let $\varphi_0 \colon G \to A[G]$ be the map given by $\varphi_0(x) = 1 \cdot x$. It is immediately verified that φ_0 is a multiplicative monoid-homomorphism, and is in fact injective, i.e. an embedding.

Let $f_0 \colon A \to A[G]$ be the map given by

$$f_0(a) = a \cdot e.$$

It is immediately verified that f_0 is a ring-homomorphism, and is also an embedding. Thus we view A as a subring of $A[G]$. One calls $A[G]$ the **monoid ring** or **monoid algebra** of G over A, or the **group algebra** if G is a group.

Examples. When G is a finite group and $A = k$ is a field, then the group ring $k[G]$ will be studied in Chapter XVIII.

Polynomial rings are special cases of the above construction. In n variables, consider a multiplicative free abelian group of rank n. Let $X_1, \dots, X_n$ be generators. Let G be the multiplicative subset consisting of elements $X_1^{\nu_1} \cdots X_n^{\nu_n}$ with $\nu_i \geqq 0$ for all i. Then G is a monoid, and the reader can verify at once that $A[G]$ is just $A[X_1, \dots, X_n]$.

As a matter of notation we usually omit the dot in writing an element of the ring $A[G]$, so we write simply $\sum a_x x$ for such an element.

More generally, let $I = \{i\}$ be an infinite family of indices, and let S be the free abelian group with free generators X_i, written multiplicatively. Then we can form the polynomial ring $A[X]$ by taking the monoid to consist of products

$$M_{(\nu)}(X) = \prod_{i \in I} X_i^{\nu_i},$$

where of course all but a finite number of exponents ν_i are equal to 0. If A is a subring of the commutative ring B, and S is a subset of B, then we shall also use the following notation. Let $\nu: S \to \mathbf{N}$ be a mapping which is 0 except for a finite number of elements of S. We write

$$M_{(\nu)}(S) = \prod_{x \in S} x^{\nu(x)}.$$

Thus we get polynomials in infinitely many variables. One interesting example of the use of such polynomials will occur in Artin's proof of the existence of the algebraic closure of a field, cf. Chapter V, Theorem 2.5.

We now consider the evaluation and reduction homomorphisms in the present context of monoids.

Proposition 3.1. *Let $\varphi: G \to G'$ be a homomorphism of monoids. Then there exists a unique homomorphism $h: A[G] \to A[G']$ such that $h(x) = \varphi(x)$ for all $x \in G$ and $h(a) = a$ for all $a \in A$.*

Proof. In fact, let $\alpha = \sum a_x x \in A[G]$. Define

$$h(\alpha) = \sum a_x \varphi(x).$$

Then h is immediately verified to be a homomorphism of abelian groups, and $h(x) = \varphi(x)$. Let $\beta = \sum b_y y$. Then

$$h(\alpha\beta) = \sum_z \left(\sum_{xy=z} a_x b_y \right) \varphi(z).$$

We get $h(\alpha\beta) = h(\alpha)h(\beta)$ immediately from the hypothesis that $\varphi(xy) =$

$\varphi(x)\varphi(y)$. If e is the unit element of G, then by definition $\varphi(e) = e'$, so Proposition 3.1 follows.

Proposition 3.2. *Let G be a monoid and let $f: A \to B$ be a homomorphism of commutative rings. Then there is a unique homomorphism*

$$h: A[G] \to B[G]$$

such that

$$h\left(\sum_{x \in G} a_x x\right) = \sum_{x \in G} f(a_x)x.$$

Proof. Since every element of $A[G]$ has a unique expression as a sum $\sum a_x x$, the formula giving h gives a well-defined map from $A[G]$ into $B[G]$. This map is obviously a homomorphism of abelian groups. As for multiplication, let

$$\alpha = \sum a_x x \qquad \text{and} \qquad \beta = \sum b_y y.$$

Then

$$\begin{aligned} h(\alpha\beta) &= \sum_{z \in G} f\left(\sum_{xy=z} a_x b_y\right) z \\ &= \sum_{z \in G} \sum_{xy=z} f(a_x) f(b_y) z \\ &= f(\alpha) f(\beta). \end{aligned}$$

We have trivially $h(1) = 1$, so h is a ring-homomorphism, as was to be shown.

Observe that viewing A as a subring of $A[G]$, the restriction of h to A is the homomorphism f itself. In other words, if e is the unit element of G, then

$$h(ae) = f(a)e.$$

§4. LOCALIZATION

We continue to let A be a commutative ring.

By a **multiplicative subset** of A we shall mean a submonoid of A (viewed as a multiplicative monoid according to **RI 2**). In other words, it is a subset S containing 1, and such that, if $x, y \in S$, then $xy \in S$.

We shall now construct the **quotient ring of** A **by** S, also known as the **ring of fractions of** A **by** S.

We consider pairs (a, s) with $a \in A$ and $s \in S$. We define a relation

$$(a, s) \sim (a', s')$$

between such pairs, by the condition that there exists an element $s_1 \in S$ such

that

$$s_1(s'a - sa') = 0.$$

It is then trivially verified that this is an equivalence relation, and the equivalence class containing a pair (a, s) is denoted by a/s. The set of equivalence classes is denoted by $S^{-1}A$.

Note that if $0 \in S$, then $S^{-1}A$ has precisely one element, namely $0/1$.

We define a multiplication in $S^{-1}A$ by the rule

$$(a/s)(a'/s') = aa'/ss'.$$

It is trivially verified that this is well defined. This multiplication has a unit element, namely $1/1$, and is clearly associative.

We define an addition in $S^{-1}A$ by the rule

$$\frac{a}{s} + \frac{a'}{s'} = \frac{s'a + sa'}{ss'}.$$

It is trivially verified that this is well defined. As an example, we give the proof in detail. Let $a_1/s_1 = a/s$, and let $a_1'/s_1' = a'/s'$. We must show that

$$(s_1'a_1 + s_1a_1')/s_1s_1' = (s'a + sa')/ss'.$$

There exist $s_2, s_3 \in S$ such that

$$s_2(sa_1 - s_1a) = 0,$$

$$s_3(s'a_1' - s_1'a') = 0.$$

We multiply the first equation by $s_3s's_1'$ and the second by s_2ss_1. We then add, and obtain

$$s_2s_3[s's_1'(sa_1 - s_1a) + ss_1(s'a_1' - s_1'a')] = 0.$$

By definition, this amounts to what we want to show, namely that there exists an element of S (e.g. s_2s_3) which when multiplied with

$$ss'(s_1'a_1 + s_1a_1') - s_1s_1'(s'a + sa')$$

yields 0.

We observe that given $a \in A$ and $s, s' \in S$ we have

$$a/s = s'a/s's.$$

Thus this aspect of the elementary properties of fractions still remains true in our present general context.

Finally, it is also trivially verified that our two laws of composition on $S^{-1}A$ define a ring structure.

We let

$$\varphi_S \colon A \to S^{-1}A$$

be the map such that $\varphi_S(a) = a/1$. Then one sees at once that φ_S is a

ring-homomorphism. Furthermore, every element of $\varphi_S(S)$ is invertible in $S^{-1}A$ (the inverse of $s/1$ is $1/s$).

Let $\mathcal{C}$ be the category whose objects are ring-homomorphisms

$$f: A \to B$$

such that for every $s \in S$, the element $f(s)$ is invertible in B. If $f: A \to B$ and $f': A \to B'$ are two objects of $\mathcal{C}$, a morphism g of f into f' is a homomorphism

$$g: B \to B'$$

making the diagram commutative:

$$\begin{array}{ccc} A & \xrightarrow{f} & B \\ & {\scriptstyle f'}\searrow & \swarrow{\scriptstyle g} \\ & B' & \end{array}$$

We contend that φ_S is a universal object in this category $\mathcal{C}$.

Proof. Suppose that $a/s = a'/s'$, or in other words that the pairs (a, s) and (a', s') are equivalent. There exists $s_1 \in S$ such that

$$s_1(s'a - sa') = 0.$$

Let $f: A \to B$ be an object of $\mathcal{C}$. Then

$$f(s_1)[f(s')f(a) - f(s)f(a')] = 0.$$

Multiplying by $f(s_1)^{-1}$, and then by $f(s')^{-1}$ and $f(s)^{-1}$, we obtain

$$f(a)f(s)^{-1} = f(a')f(s')^{-1}.$$

Consequently, we can define a map

$$h: S^{-1}A \to B$$

such that $h(a/s) = f(a)f(s)^{-1}$, for all $a/s \in S^{-1}A$. It is trivially verified that h is a homomorphism, and makes the usual diagram commutative. It is also trivially verified that such a map h is unique, and hence that φ_S is the required universal object.

Let A be an entire ring, and let S be a multiplicative subset which does not contain 0. Then

$$\varphi_S: A \to S^{-1}A$$

is injective.

Indeed, by definition, if $a/1 = 0$ then there exists $s \in S$ such that $sa = 0$, and hence $a = 0$.

The most important cases of a multiplicative set S are the following:

1. Let A be a commutative ring, and let S be the set of invertible elements of A (i.e. the set of units). Then S is obviously multiplicative, and is

denoted frequently by A^*. If A is a field, then A^* is the multiplicative group of non-zero elements of A. In that case, $S^{-1}A$ is simply A itself.

2. Let A be an entire ring, and let S be the set of non-zero elements of A. Then S is a multiplicative set, and $S^{-1}A$ is then a field, called the **quotient field** or the **field of fractions**, of A. It is then customary to identify A as a subset of $S^{-1}A$, and we can write

$$a/s = s^{-1}a$$

for $a \in A$ and $s \in S$.

We have seen in §3 that when A is an entire ring, then $A[X_1, \dots, X_n]$ is also entire. If K is the quotient field of A, the quotient field of $A[X_1, \dots, X_n]$ is denoted by $K(X_1, \dots, X_n)$. An element of $K(X_1, \dots, X_n)$ is called a **rational function**. A rational function can be written as a quotient $f(X)/g(X)$ where f, g are polynomials. If $(b_1, \dots, b_n)$ is in $K^{(n)}$, and a rational function admits an expression as a quotient f/g such that $g(b) \neq 0$, then we say that the rational function is **defined** at (b). From general localization properties, we see that when this is the case, we can substitute (b) in the rational function to get a value $f(b)/g(b)$.

3. A ring A is called a **local ring** if it is commutative and has a unique maximal ideal. If A is a local ring and $\mathfrak{m}$ is its maximal ideal, and $x \in A$, $x \notin \mathfrak{m}$, then x is a unit (otherwise x generates a proper ideal, not contained in $\mathfrak{m}$, which is impossible). Let A be a ring and $\mathfrak{p}$ a prime ideal. Let S be the complement of $\mathfrak{p}$ in A. Then S is a multiplicative subset of A, and $S^{-1}A$ is denoted by $A_\mathfrak{p}$. It is a local ring (cf. Exercise 3) and is called **the local ring of** A **at** $\mathfrak{p}$. Cf. the examples of principal rings, and Exercises 15, 16.

Let S be a multiplicative subset of A. Denote by $J(A)$ the set of ideals of A. Then we can define a map

$$\psi_S: J(A) \to J(S^{-1}A);$$

namely we let $\psi_S(\mathfrak{a}) = S^{-1}\mathfrak{a}$ be the subset of $S^{-1}A$ consisting of all fractions a/s with $a \in \mathfrak{a}$ and $s \in S$. The reader will easily verify that $S^{-1}\mathfrak{a}$ is an $S^{-1}A$-ideal, and that ψ_S is a homomorphism for both the additive and multiplicative monoid structures on the set of ideals $J(A)$. Furthermore, ψ_S also preserves intersections and inclusions; in other words, for ideals $\mathfrak{a}$, $\mathfrak{b}$ of A we have:

$$S^{-1}(\mathfrak{a} + \mathfrak{b}) = S^{-1}\mathfrak{a} + S^{-1}\mathfrak{b}, \qquad S^{-1}(\mathfrak{a}\mathfrak{b}) = (S^{-1}\mathfrak{a})(S^{-1}\mathfrak{b}),$$

$$S^{-1}(\mathfrak{a} \cap \mathfrak{b}) = S^{-1}\mathfrak{a} \cap S^{-1}\mathfrak{b}.$$

As an example, we prove this last relation. Let $x \in \mathfrak{a} \cap \mathfrak{b}$. Then x/s is in $S^{-1}\mathfrak{a}$ and also in $S^{-1}\mathfrak{b}$, so the inclusion is trivial. Conversely, suppose we have an element of $S^{-1}A$ which can be written as $a/s = b/s'$ with $a \in \mathfrak{a}$, $b \in \mathfrak{b}$, and $s, s' \in S$. Then there exists $s_1 \in S$ such that

$$s_1 s' a = s_1 s b,$$

and this element lies in both $\mathfrak{a}$ and $\mathfrak{b}$. Hence

$$a/s = s_1 s'a/s_1 s's$$

lies in $S^{-1}(\mathfrak{a} \cap \mathfrak{b})$, as was to be shown.

§5. PRINCIPAL AND FACTORIAL RINGS

Let A be an entire ring. An element $a \neq 0$ is called **irreducible** if it is not a unit, and if whenever one can write $a = bc$ with $b \in A$ and $c \in A$ then b or c is a unit.

Let $a \neq 0$ be an element of A and assume that the principal ideal (a) is prime. Then a is irreducible. Indeed, if we write $a = bc$, then b or c lies in (a), say b. Then we can write $b = ad$ with some $d \in A$, and hence $a = acd$. Since A is entire, it follows that $cd = 1$, in other words, that c is a unit.

The converse of the preceding assertion is not always true. We shall discuss under which conditions it is true. An element $a \in A$, $a \neq 0$, is said to have a **unique factorization into irreducible elements** if there exists a unit u and there exist irreducible elements p_i $(i = 1, \dots, r)$ in A such that

$$a = u \prod_{i=1}^{r} p_i,$$

and if given two factorizations into irreducible elements,

$$a = u \prod_{i=1}^{r} p_i = u' \prod_{j=1}^{s} q_j,$$

we have $r = s$, and after a permutation of the indices i, we have $p_i = u_i q_i$ for some unit u_i in A, $i = 1, \dots, r$.

We note that if p is irreducible and u is a unit, then up is also irreducible, so we must allow multiplication by units in a factorization. In the ring of integers $\mathbf{Z}$, the ordering allows us to select a representative irreducible element (a prime number) out of two possible ones differing by a unit, namely $\pm p$, by selecting the positive one. This is, of course, impossible in more general rings.

Taking $r = 0$ above, we adopt the convention that a unit of A has a factorization into irreducible elements.

A ring is called **factorial** (or a **unique factorization ring**) if it is entire and if every element $\neq 0$ has a unique factorization into irreducible elements. We shall prove below that a principal entire ring is factorial.

Let A be an entire ring and $a, b \in A$, $ab \neq 0$. We say that a **divides** b and write $a|b$ if there exists $c \in A$ such that $ac = b$. We say that $d \in A$, $d \neq 0$, is a **greatest common divisor (g.c.d.)** of a and b if $d|a$, $d|b$, and if any element e of A, $e \neq 0$, which divides both a and b also divides d.

Proposition 5.1. *Let A be a principal entire ring and $a, b \in A$, $a, b \neq 0$. Let $(a, b) = (c)$. Then c is a greatest common divisor of a and b.*

Proof. Since b lies in the ideal (c), we can write $b = xc$ for some $x \in A$, so that $c|b$. Similarly, $c|a$. Let d divide both a and b, and write $a = dy$, $b = dz$ with $y, z \in A$. Since c lies in (a, b) we can write

$$c = wa + tb$$

with some $w, t \in A$. Then $c = w\,dy + t\,dz = d(wy + tz)$, whence $d|c$, and our proposition is proved.

Theorem 5.2. *Let A be a principal entire ring. Then A is factorial.*

Proof. We first prove that every non-zero element of A has a factorization into irreducible elements. Let S be the set of principal ideals $\neq 0$ whose generators do not have a factorization into irreducible elements, and suppose S is not empty. Let (a_1) be in S. Consider an ascending chain

$$(a_1) \subsetneqq (a_2) \subsetneqq \cdots \subsetneqq (a_n) \subsetneqq \cdots$$

of ideals in S. We contend that such a chain cannot be infinite. Indeed, the union of such a chain is an ideal of A, which is principal, say equal to (a). The generator a must already lie in some element of the chain, say (a_n), and then we see that $(a_n) \subset (a) \subset (a_n)$, whence the chain stops at (a_n). Hence S is inductively ordered, and has a maximal element (a). Therefore any ideal of A containing (a) and $\neq (a)$ has a generator admitting a factorization.

We note that a_n cannot be irreducible (otherwise it has a factorization), and hence we can write $a = bc$ with neither b nor c equal to a unit. But then $(b) \neq (a)$ and $(c) \neq (a)$ and hence both b, c admit factorizations into irreducible elements. The product of these factorizations is a factorization for a, contradicting the assumption that S is not empty.

To prove uniqueness, we first remark that if p is an irreducible element of A and $a, b \in A$, $p|ab$, then $p|a$ or $p|b$. *Proof*: If $p \nmid a$, then the g.c.d. of p, a is 1 and hence we can write

$$1 = xp + ya$$

with some $x, y \in A$. Then $b = bxp + yab$, and since $p|ab$ we conclude that $p|b$.

Suppose that a has two factorizations

$$a = p_1 \cdots p_r = q_1 \cdots q_s$$

into irreducible elements. Since p_1 divides the product farthest to the right, p_1 divides one of the factors, which we may assume to be q_1 after renumbering these factors. Then there exists a unit u_1 such that $q_1 = u_1 p_1$. We can now cancel p_1 from both factorizations and get

$$p_2 \cdots p_r = u_1 q_2 \cdots q_s.$$

The argument is completed by induction.

We could call two elements a, $b \in A$ equivalent if there exists a unit u such that $a = bu$. Let us select one irreducible element p out of each equivalence class belonging to such an irreducible element, and let us denote by P the set of such representatives. Let $a \in A$, $a \neq 0$. Then there exists a unit u and integers $v(p) \geqq 0$, equal to 0 for almost all $p \in P$ such that

$$a = u \prod_{p \in P} p^{v(p)}.$$

Furthermore, the unit u and the integers $v(p)$ are uniquely determined by a. We call $v(p)$ the **order** of a at p, also written $\text{ord}_p\, a$.

If A is a factorial ring, then an irreducible element p generates a prime ideal (p). Thus in a factorial ring, an irreducible element will also be called a **prime element**, or simply a **prime**.

We observe that one can define the notion of **least common multiple** (l.c.m.) of a finite number of non-zero elements of A in the usual manner: If

$$a_1, \ldots, a_n \in A$$

are such elements, we define a l.c.m. for these elements to be any $c \in A$ such that for all primes p of A we have

$$\text{ord}_p\, c = \max_i \text{ord}_p\, a_i.$$

This element c is well defined up to a unit.

If a, $b \in A$ are non-zero elements, we say that a, b are **relaively prime** if the g.c.d. of a and b is a unit.

Example. The ring of integers **Z** is factorial. Its group of units consists of 1 and -1. It is natural to take as representative prime element the positive prime element (what is called a prime number) p from the two possible choices p and $-p$. Similarly, we shall show later that the ring of polynomials in one variable over a field is factorial, and one selects representatives for the prime elements to be the irreducible polynomials with leading coefficient 1.

Examples. It will be proved in Chapter IV that if R is a factorial ring, then the polynomial ring $R[X_1, \ldots, X_n]$ in n variables is factorial. In particular, if k is a field, then the polynomial ring $k[X_1, \ldots, X_n]$ is factorial. Note that $k[X_1]$ is a principal ring, but for $n \geqq 2$, the ring $k[X_1, \ldots, X_n]$ is not principal.

In Exercise 5 you will prove that the localization of a factorial ring is factorial.

In Chapter IV, §9 we shall prove that the power series ring $k[[X_1, \ldots, X_n]]$ is factorial. This result is a special case of the more general statement that a regular local ring is factorial, but we do not define regular local rings in this book. You can look them up in books on commutative

algebra. I recommend:

H. MATSUMURA, *Commutative Algebra*, second edition, Benjamin-Cummings, New York, 1980

H. MATSUMURA, *Commutative Rings*, Cambridge University Press, Cambridge, UK, 1986

Examples from algebraic and complex geometry. Roughly speaking, regular local rings arise in the following context of algebraic or complex geometry. Consider the ring of regular functions in the neighborhood of some point on a complex or algebraic manifold. This ring is regular. A typical example is the ring of convergent power series in a neighborhood of 0 in $\mathbf{C}^n$. In Chapter IV, we shall prove some results on power series which give some algebraic background for those analytic theories, and which are used in proving the factoriality of rings of power series, convergent or not.

Conversely to the above examples, singularities in geometric theories may give rise to examples of non-factoriality. We give examples using notions which are sufficiently basic so that readers should have encountered them in more elementary courses.

Examples of non-factorial rings. Let k be a field, and let x be a variable over k. Let $R = k[x^2, x^3]$. Then R is not factorial (proof?). The ring R may be viewed as the ring of regular functions on the curve $y^2 = x^3$, which has a singularity at the origin, as you can see by drawing its real graph.

Let R be the set of all numbers of the form $a + b\sqrt{-5}$, where $a, b \in \mathbf{Z}$. Then the only units of R are ± 1, and the elements $3, 2 + \sqrt{-5}, 2 - \sqrt{-5}$ are irreducible elements, giving rise to a non-unique factorization

$$3^2 = (2 + \sqrt{-5})(2 - \sqrt{-5}).$$

(Do Exercise 10.) Here the non-factoriality is not due to singularities but due to a non-trivial ideal class group of R, which is a Dedekind ring. For a definition see the exercises of Chapter III, or go straight to my book *Algebraic Number Theory*, for instance.

As Trotter once pointed out (*Math. Monthly*, April 1988), the relation

$$\sin^2 x = (1 + \cos x)(1 - \cos x)$$

may be viewed as a non-unique factorization in the ring of trigonometric polynomials $\mathbf{R}[\sin x, \cos x]$, generated over $\mathbf{R}$ by the functions $\sin x$ and $\cos x$. This ring is a subring of the ring of all functions, or of all differentiable functions. See Exercise 11.

EXERCISES

We let A denote a commutative ring.

1. Suppose that $1 \neq 0$ in A. Let S be a multiplicative subset of A not containing 0. Let $\mathfrak{p}$ be a maximal element in the set of ideals of A whose intersection with S is empty. Show that $\mathfrak{p}$ is prime.

2. Let $f: A \to A'$ be a surjective homomorphism of rings, and assume that A is local, $A' \neq 0$. Show that A' is local.

3. Let $\mathfrak{p}$ be a prime ideal of A. Show that $A_{\mathfrak{p}}$ has a unique maximal ideal, consisting of all elements a/s with $a \in \mathfrak{p}$ and $s \notin \mathfrak{p}$.

4. Let A be a principal ring and S a multiplicative subset with $0 \notin S$. Show that $S^{-1}A$ is principal.

5. Let A be a factorial ring and S a multiplicative subset with $0 \notin S$. Show that $S^{-1}A$ is factorial, and that the prime elements of $S^{-1}A$ are those primes p of A such that $(p) \cap S$ is empty.

6. Let A be a factorial ring and p a prime element. Show that the local ring $A_{(p)}$ is principal.

7. Let A be a principal ring and $a_1, \ldots, a_n$ non-zero elements of A. Let $(a_1, \ldots, a_n) = (d)$. Show that d is a greatest common divisor for the a_i $(i = 1, \ldots, n)$.

8. Let p be a prime number, and let A be the ring $\mathbf{Z}/p^r\mathbf{Z}$ (r = integer $\geqq 1$). Let G be the group of units in A, i.e. the group of integers prime to p, modulo p^r. Show that G is cyclic, except in the case when

$$p = 2, \qquad r \geqq 3,$$

in which case it is of type $(2, 2^{r-2})$. [*Hint*: In the general case, show that G is the product of a cyclic group generated by $1 + p$, and a cyclic group of order $p - 1$. In the exceptional case, show that G is the product of the group $\{\pm 1\}$ with the cyclic group generated by the residue class of 5 mod 2^r.]

9. Let i be the complex number $\sqrt{-1}$. Show that the ring $\mathbf{Z}[i]$ is principal, and hence factorial. What are the units?

10. Let D be an integer $\geqq 1$, and let R be the set of all element $a + b\sqrt{-D}$ with $a, b \in \mathbf{Z}$.
    (a) Show that R is a ring.
    (b) Using the fact that complex conjugation is an automorphism of $\mathbf{C}$, show that complex conjugation induces an automorphism of R.
    (c) Show that if $D \geqq 2$ then the only units in R are ± 1.
    (d) Show that $3, 2 + \sqrt{-5}, 2 - \sqrt{-5}$ are irreducible elements in $\mathbf{Z}[\sqrt{-5}]$.

11. Let R be the ring of trigonometric polynomials as defined in the text. Show that R consists of all functions f on $\mathbf{R}$ which have an expression of the form

$$f(x) = a_0 + \sum_{m=1}^{n} (a_m \cos mx + b_m \sin mx),$$

where a_0, a_m, b_m are real numbers. Define the **trigonometric degree** $\deg_{\text{tr}}(f)$ to be the maximum of the integers r, s such that $a_r, b_s \neq 0$. Prove that

$$\deg_{\text{tr}}(fg) = \deg_{\text{tr}}(f) + \deg_{\text{tr}}(g).$$

Deduce from this that R has no divisors of 0, and also deduce that the functions $\sin x$ and $1 - \cos x$ are irreducible elements in that ring.

12. Let P be the set of positive integers and R the set of functions defined on P with values in a commutative ring K. Define the sum in R to be the ordinary addition of functions, and define the **convolution product** by the formula

$$(f * g)(m) = \sum_{xy=m} f(x)g(y),$$

where the sum is taken over all pairs (x, y) of positive integers such that $xy = m$.
   (a) Show that R is a commutative ring, whose unit element is the function δ such that $\delta(1) = 1$ and $\delta(x) = 0$ if $x \neq 1$.
   (b) A function f is said to be **multiplicative** if $f(mn) = f(m)f(n)$ whenever m, n are relatively prime. If f, g are multiplicative, show that $f * g$ is multiplicative.
   (c) Let μ be the **Möbius function** such that $\mu(1) = 1$, $\mu(p_1 \cdots p_r) = (-1)^r$ if $p_1, \ldots, p_r$ are distinct primes, and $\mu(m) = 0$ if m is divisible by p^2 for some prime p. Show that $\mu * \varphi_1 = \delta$, where φ_1 denotes the constant function having value 1. [*Hint*: Show first that μ is multiplicative, and then prove the assertion for prime powers.] The Möbius inversion formula of elementary number theory is then nothing else but the relation $\mu * \varphi_1 * f = f$.

Dedekind rings

Prove the following statements about a Dedekind ring $\mathfrak{o}$. To simplify terminology, by an **ideal** we shall mean non-zero ideal unless otherwise specified. We let K denote the quotient field of $\mathfrak{o}$.

13. Every ideal is finitely generated. [*Hint*: Given an ideal $\mathfrak{a}$, let $\mathfrak{b}$ be the fractional ideal such that $\mathfrak{a}\mathfrak{b} = \mathfrak{o}$. Write $1 = \sum a_i b_i$ with $a_i \in \mathfrak{a}$ and $b_i \in \mathfrak{b}$. Show that $\mathfrak{a} = (a_1, \ldots, a_n)$.]

14. Every ideal has a factorization as a product of prime ideals, uniquely determined up to permutation.

15. Suppose $\mathfrak{o}$ has only one prime ideal $\mathfrak{p}$. Let $t \in \mathfrak{p}$ and $t \notin \mathfrak{p}^2$. Then $\mathfrak{p} = (t)$ is principal.

16. Let $\mathfrak{o}$ be any Dedekind ring. Let $\mathfrak{p}$ be a prime ideal. Let $\mathfrak{o}_{\mathfrak{p}}$ be the local ring at $\mathfrak{p}$. Then $\mathfrak{o}_{\mathfrak{p}}$ is Dedekind and has only one prime ideal.

17. As for the integers, we say that $\mathfrak{a}|\mathfrak{b}$ ($\mathfrak{a}$ **divides** $\mathfrak{b}$) if there exists an ideal $\mathfrak{c}$ such that $\mathfrak{b} = \mathfrak{a}\mathfrak{c}$. Prove:
   (a) $\mathfrak{a}|\mathfrak{b}$ if and only if $\mathfrak{b} \subset \mathfrak{a}$.
   (b) Let $\mathfrak{a}$, $\mathfrak{b}$ be ideals. Then $\mathfrak{a} + \mathfrak{b}$ is their greatest common divisor. In particular, $\mathfrak{a}$, $\mathfrak{b}$ are relatively prime if and only if $\mathfrak{a} + \mathfrak{b} = \mathfrak{o}$.

18. Every prime ideal $\mathfrak{p}$ is maximal. (Remember, $\mathfrak{p} \neq 0$ by convention.) In particular, if $\mathfrak{p}_1, \ldots, \mathfrak{p}_n$ are distinct primes, then the Chinese remainder theorem applies to their powers $\mathfrak{p}_1^{r_1}, \ldots, \mathfrak{p}_n^{r_n}$. Use this to prove:

19. Let $\mathfrak{a}$, $\mathfrak{b}$ be ideals. Show that there exists an element $c \in K$ (the quotient field of $\mathfrak{o}$) such that $c\mathfrak{a}$ is an ideal relatively prime to $\mathfrak{b}$. In particular, every ideal class in $\mathrm{Pic}(\mathfrak{o})$ contains representative ideals prime to a given ideal.
    For a continuation, see Exercise 7 of Chapter VII.

CHAPTER III

Modules

Although this chapter is logically self-contained and prepares for future topics, in practice readers will have had some acquaintance with vector spaces over a field. We generalize this notion here to modules over rings. It is a standard fact (to be reproved) that a vector space has a basis, but for modules this is not always the case. Sometimes they do; most often they do not. We shall look into cases where they do.

For examples of modules and their relations to those which have a basis, the reader should look at the comments made at the end of §4.

§1. BASIC DEFINITIONS

Let A be a ring. A **left module** over A, or a left A-module M is an abelian group, usually written additively, together with an operation of A on M (viewing A as a multiplicative monoid by **RI 2**), such that, for all $a, b \in A$ and $x, y \in M$ we have

$$(a + b)x = ax + bx \quad \text{and} \quad a(x + y) = ax + ay.$$

We leave it as an exercise to prove that $a(-x) = -(ax)$ and that $0x = 0$. By definition of an operation, we have $1x = x$.

In a similar way, one defines a **right A-module**. We shall deal only with left A-modules, unless otherwise specified, and hence call these simply **A-modules**, or even **modules** if the reference is clear.

Let M be an A-module. By a **submodule** N of M we mean an additive subgroup such that $AN \subset N$. Then N is a module (with the operation induced by that of A on M).

Examples

We note that A is a module over itself.

Any commutative group is a $\mathbf{Z}$-module.

An additive group consisting of 0 alone is a module over any ring.

Any left ideal of A is a module over A.

Let J be a two-sided ideal of A. Then the factor ring A/J is actually a module over A. If $a \in A$ and $a + J$ is a coset of J in A, then one defines the operation to be $a(x + J) = ax + J$. The reader can verify at once that this defines a module structure on A/J. More general, if M is a module and N a submodule, we shall define the factor module below. Thus if L is a left ideal of A, then A/L is also a module. For more examples in this vein, see §4.

A module over a field is called a **vector space**. Even starting with vector spaces, one is led to consider modules over rings. Indeed, let V be a vector space over the field K. The reader no doubt already knows about linear maps (which will be recalled below systematically). Let R be the ring of all linear maps of V into itself. Then V is a module over R. Similarly, if $V = K^n$ denotes the vector space of (vertical) n-tuples of elements of K, and R is the ring of $n \times n$ matrices with components in K, then V is a module over R. For more comments along these lines, see the examples at the end of §2.

Let S be a non-empty set and M an A-module. Then the set of maps $\mathrm{Map}(S, M)$ is an A-module. We have already noted previously that it is a commutative group, and for $f \in \mathrm{Map}(S, M)$, $a \in A$ we define af to be the map such that $(af)(s) = af(s)$. The axioms for a module are then trivially verified.

For further examples, see the end of this section.

For the rest of this section, we deal with a fixed ring A, and hence may omit the prefix A-.

Let A be an *entire* ring and let M be an A-module. We define the **torsion submodule** M_{tor} to be the subset of elements $x \in M$ such that there exists $a \in A$, $a \neq 0$ such that $ax = 0$. It is immediately verified that M_{tor} is a submodule. Its structure in an important case will be determined in §7.

Let $\mathfrak{a}$ be a left ideal, and M a module. We define $\mathfrak{a}M$ to be the set of all elements

$$a_1x_1 + \cdots + a_nx_n$$

with $a_i \in \mathfrak{a}$ and $x_i \in M$. It is obviously a submodule of M. If $\mathfrak{a}, \mathfrak{b}$ are left ideals, then we have associativity, namely

$$\mathfrak{a}(\mathfrak{b}M) = (\mathfrak{a}\mathfrak{b})M.$$

We also have some obvious distributivities, like $(\mathfrak{a} + \mathfrak{b})M = \mathfrak{a}M + \mathfrak{b}M$. If N, N' are submodules of M, then $\mathfrak{a}(N + N') = \mathfrak{a}N + \mathfrak{a}N'$.

Let M be an A-module, and N a submodule. We shall define a module structure on the factor group M/N (for the additive group structure). Let $x + N$ be a coset of N in M, and let $a \in A$. We define $a(x + N)$ to be the coset $ax + N$. It is trivial to verify that this is well defined (i.e. if y is in the same coset as x, then ay is in the same coset as ax), and that this is an operation of A on M/N satisfying the required condition, making M/N into a module, called the **factor module** of M by N.

By a **module-homomorphism** one means a map

$$f: M \to M'$$

of one module into another (over the same ring A), which is an additive group-homomorphism, and such that

$$f(ax) = af(x)$$

for all $a \in A$ and $x \in M$. It is then clear that the collection of A-modules is a category, whose morphisms are the module-homomorphisms usually also called homomorphisms for simplicity, if no confusion is possible. If we wish to refer to the ring A, we also say that f is an A-**homomorphism**, or also that it is an A-**linear map**.

If M is a module, then the identity map is a homomorphism. For any module M', the map $\zeta: M \to M'$ such that $\zeta(x) = 0$ for all $x \in M$ is a homomorphism, called **zero**.

In the next section, we shall discuss the homomorphisms of a module into itself, and as a result we shall give further examples of modules which arise in practice. Here we continue to tabulate the translation of basic properties of groups to modules.

Let M be a module and N a submodule. We have the canonical additive group-homomorphism

$$f: M \to M/N$$

and one verifies trivially that it is a module-homomorphism.

Equally trivially, one verifies that f is universal in the category of homomorphisms of M whose kernel contains N.

If $f: M \to M'$ is a module-homomorphism, then its kernel and image are submodules of M and M' respectively (trivial verification).

Let $f: M \to M'$ be a homomorphism. By the **cokernel** of f we mean the factor module $M'/\text{Im} f = M'/f(M)$. One may also mean the canonical homomorphism

$M' \to M'/f(M)$ rather than the module itself. The context should make clear which is meant. Thus the cokernel is a factor module of M'.

Canonical homomorphisms discussed in Chapter I, §3 apply to modules *mutatis mutandis*. For the convenience of the reader, we summarise these homomorphisms:

Let N, N' be two submodules of a module M. Then $N + N'$ is also a submodule, and we have an isomorphism

$$N/(N \cap N') \approx (N + N')/N'.$$

If $M \supset M' \supset M''$ are modules, then

$$(M/M'')/(M'/M'') \approx M/M'.$$

If $f: M \to M'$ is a module-homomorphism, and N' is a submodule of M', then $f^{-1}(N')$ is a submodule of M and we have a canonical injective homomorphism

$$\bar{f}: M/f^{-1}(N') \to M'/N'.$$

If f is surjective, then $\bar{f}$ is a module-isomorphism.

The proofs are obtained by verifying that all homomorphisms which appeared when dealing with abelian groups are now A-homomorphisms of modules. We leave the verification to the reader.

As with groups, we observe that a module-homomorphism which is bijective is a module-isomorphism. Here again, the proof is the same as for groups, adding only the observation that the inverse map, which we know is a group-isomorphism, actually is a module-isomorphism. Again, we leave the verification to the reader.

As with abelian groups, we define a sequence of module-homomorphisms

$$M' \xrightarrow{f} M \xrightarrow{g} M''$$

to be **exact** if $\operatorname{Im} f = \operatorname{Ker} g$. We have an exact sequence associated with a submodule N of a module M, namely

$$0 \to N \to M \to M/N \to 0,$$

the map of N into M being the inclusion, and the subsequent map being the canonical map. The notion of exactness is due to Eilenberg-Steenrod.

If a homomorphism $u: N \to M$ is such that

$$0 \to N \xrightarrow{u} M$$

is exact, then we also say that u is a **monomorphism** or an **embedding**. Dually, if

$$N \xrightarrow{u} M \to 0$$

is exact, we say that u is an **epimorphism**.

Algebras

There are some things in mathematics which satisfy all the axioms of a ring except for the existence of a unit element. We gave the example of $L^1(\mathbf{R})$ in Chapter II, §1. There are also some things which do not satisfy associativity, but satisfy distributivity. For instance let R be a ring, and for $x, y \in R$ define the **bracket product**

$$[x, y] = xy - yx.$$

Then this bracket product is not associative in most cases when R is not commutative, but it satisfies the distributive law.

Examples. A typical example is the ring of differential operators with C^∞ coefficients, operating on the ring of C^∞ functions on an open set in $\mathbf{R}^n$. The bracket product

$$[D_1, D_2] = D_1 \circ D_2 - D_2 \circ D_1$$

of two differential operators is again a differential operator. In the theory of Lie groups, the tangent space at the origin also has such a bracket product.

Such considerations lead us to define a more general notion than a ring. Let A be a commutative ring. Let E, F be modules. By a **bilinear map**

$$g: E \times E \to F$$

we mean a map such that given $x \in E$, the map $y \mapsto g(x, y)$ is A-linear, and given $y \in E$, the map $x \mapsto g(x, y)$ is A-linear. By an A-**algebra** we mean a module together with a bilinear map $g: E \times E \to E$. We view such a map as a law of composition on E. But in this book, unless otherwise specified, we shall assume that our algebras are associative and have a unit element.

Aside from the examples already mentioned, we note that the group ring $A[G]$ (or monoid ring when G is a monoid) is an A-algebra, also called the **group** (or **monoid**) **algebra**. Actually the group algebra can be viewed as a special case of the following situation.

Let $f: A \to B$ be a ring-homomorphism such that $f(A)$ is contained in the center of B, i.e., $f(a)$ commutes with every element of B for every $a \in A$. Then we may view B as an A-module, defining the operation of A on B by the map

$$(a, b) \mapsto f(a)b$$

for all $a \in A$ and $b \in B$. The axioms for a module are trivially satisfied, and the multiplicative law of composition $B \times B \to B$ is clearly bilinear (i.e., A-bilinear). In this book, unless otherwise specified, by an **algebra** over A, we shall always mean a ring-homomorphism as above. We say that the algebra is **finitely generated** if B is finitely generated as a ring over $f(A)$.

Several examples of modules over a polynomial algebra or a group algebra will be given in the next section, where we also establish the language of representations.

§2. THE GROUP OF HOMOMORPHISMS

Let A be a ring, and let X, X' be A-modules. We denote by $\mathrm{Hom}_A(X', X)$ the set of A-homomorphisms of X' into X. Then $\mathrm{Hom}_A(X', X)$ is an abelian group, the law of addition being that of addition for mappings into an abelian group.

If A *is commutative* then we can make $\mathrm{Hom}_A(X', X)$ into an A-module, by defining af for $a \in A$ and $f \in \mathrm{Hom}_A(X', X)$ to be the map such that

$$(af)(x) = af(x).$$

The verification that the axioms for an A-module are satisfied is trivial. However, if A is not commutative, then we view $\mathrm{Hom}_A(X', X)$ simply as an abelian group.

We also view Hom_A as a functor. It is actually a functor of two variables, contravariant in the first and covariant in the second. Indeed, let Y be an A-module, and let

$$X' \xrightarrow{f} X$$

be an A-homomorphism. Then we get an induced homomorphism

$$\mathrm{Hom}_A(f, Y) : \mathrm{Hom}_A(X, Y) \to \mathrm{Hom}_A(X', Y)$$

(reversing the arrow!) given by

$$g \mapsto g \circ f.$$

This is illustrated by the following sequence of maps:

$$X' \xrightarrow{f} X \xrightarrow{g} Y.$$

The fact that $\mathrm{Hom}_A(f, Y)$ is a homomorphism is simply a rephrasing of the property $(g_1 + g_2) \circ f = g_1 \circ f + g_2 \circ f$, which is trivially verified. If $f = \mathrm{id}$, then composition with f acts as an identity mapping on g, i.e. $g \circ \mathrm{id} = g$.

If we have a sequence of A-homomorphisms

$$X' \to X \to X'',$$

then we get an induced sequence

$$\mathrm{Hom}_A(X', Y) \leftarrow \mathrm{Hom}_A(X, Y) \leftarrow \mathrm{Hom}_A(X'', Y).$$

Proposition 2.1. *A sequence*

$$X' \xrightarrow{\lambda} X \to X'' \to 0$$

is exact if and only if the sequence

$$\mathrm{Hom}_A(X', Y) \leftarrow \mathrm{Hom}_A(X, Y) \leftarrow \mathrm{Hom}_A(X'', Y) \leftarrow 0$$

is exact for all Y.

Proof. This is an important fact, whose proof is easy. For instance, suppose the first sequence is exact. If $g : X'' \to Y$ is an A-homomorphism, its image in $\mathrm{Hom}_A(X, Y)$ is obtained by composing g with the surjective map of X on X''. If this composition is 0, it follows that $g = 0$ because $X \to X''$ is surjective. As another example, consider a homomorphism $g : X \to Y$ such that the composition

$$X' \xrightarrow{\lambda} X \xrightarrow{g} Y$$

is 0. Then g vanishes on the image of λ. Hence we can factor g through the factor module,

$$\begin{array}{ccc} & X/\mathrm{Im}\,\lambda & \\ \nearrow & & \searrow \\ X & \xrightarrow[g]{} & Y \end{array}$$

Since $X \to X''$ is surjective, we have an isomorphism

$$X/\mathrm{Im}\,\lambda \leftrightarrow X''.$$

Hence we can factor g through X'', thereby showing that the kernel of

$$\mathrm{Hom}_A(X', Y) \leftarrow \mathrm{Hom}_A(X, Y)$$

is contained in the image of

$$\mathrm{Hom}_A(X, Y) \leftarrow \mathrm{Hom}_A(X'', Y).$$

The other conditions needed to verify exactness are left to the reader. So is the converse.

We have a similar situation with respect to the second variable, but then the functor is covariant. Thus if X is fixed, and we have a sequence of A-homomorphisms

$$Y' \to Y \to Y'',$$

then we get an induced sequence

$$\mathrm{Hom}_A(X, Y') \to \mathrm{Hom}_A(X, Y) \to \mathrm{Hom}_A(X, Y'').$$

Proposition 2.2. *A sequence*

$$0 \to Y' \to Y \to Y'',$$

is exact if and only if

$$0 \to \mathrm{Hom}_A(X, Y') \to \mathrm{Hom}_A(X, Y) \to \mathrm{Hom}_A(X, Y'')$$

is exact for all X.

The verification will be left to the reader. It follows at once from the definitions.

We note that to say that

$$0 \to Y' \to Y$$

is exact means that Y' is embedded in Y, i.e. is isomorphic to a submodule of Y. A homomorphism into Y' can be viewed as a homomorphism into Y if we have $Y' \subset Y$. This corresponds to the injection

$$0 \to \operatorname{Hom}_A(X, Y') \to \operatorname{Hom}_A(X, Y).$$

Let $\operatorname{Mod}(A)$ and $\operatorname{Mod}(B)$ be the categories of modules over rings A and B, and let $F\colon \operatorname{Mod}(A) \to \operatorname{Mod}(B)$ be a functor. One says that F is **exact** if F transforms exact sequences into exact sequences. We see that the Hom functor in either variable need not be exact if the other variable is kept fixed. In a later section, we define conditions under which exactness is preserved.

Endomorphisms. Let M be an A-module. From the relations

$$(g_1 + g_2) \circ f = g_1 \circ f + g_2 \circ f$$

and its analogue on the right, namely

$$g \circ (f_1 + f_2) = g \circ f_1 + g \circ f_2,$$

and the fact that there is an identity for composition, namely id_M, we conclude that $\operatorname{Hom}_A(M, M)$ is a ring, the multiplication being defined as composition of mappings. If n is an integer $\geqq 1$, we can write f^n to mean the iteration of f with itself n times, and define f^0 to be id. According to the general definition of endomorphisms in a category, we also write $\operatorname{End}_A(M)$ instead of $\operatorname{Hom}_A(M, M)$, and we call $\operatorname{End}_A(M)$ the ring of **endomorphisms**.

Since an A-module M is an abelian group, we see that $\operatorname{Hom}_{\mathbf{Z}}(M, M)$ (= set of group-homomorphisms of M into itself) is a ring, and that we could have defined an operation of A on M to be a ring-homomorphism $A \to \operatorname{Hom}_{\mathbf{Z}}(M, M)$.

Let A be *commutative*. Then M is a module over $\operatorname{End}_A(M)$. If R is a subring of $\operatorname{End}_A(M)$ then M is *a fortiori* a module over R. More generally, let R be a ring and let $\rho\colon R \to \operatorname{End}_A(M)$ be a ring homomorphism. Then ρ is called a **representation** of R on M. This occurs especially if $A = K$ is a field. The linear algebra of representations of a ring will be discussed in Part III, in several contexts, mostly finite-dimensional. Infinite-dimensional examples occur in analysis, but then the representation theory mixes algebra with analysis, and thus goes beyond the level of this course.

Example. Let K be a field and let V be a vector space over K. Let $D\colon V \to V$ be an endomorphism (K-linear map). For every polynomial $P(X) \in K[X]$, $P(X) = \sum a_i X^i$ with $a_i \in K$, we can define

$$P(D) = \sum a_i D^i \colon V \to V$$

as an endomorphism of V. The association $P(X) \mapsto P(D)$ gives a representation

$$\rho\colon K[X] \to \mathrm{End}_K(V),$$

which makes V into a $K[X]$-module. It will be shown in Chapter IV that $K[X]$ is a principal ring. In §7 we shall give a general structure theorem for modules over principal rings, which will be applied to the above example in the context of linear algebra for finite-dimensional vector spaces in Chapter XIV, §3. Readers acquainted with basic linear algebra from an undergraduate course may wish to read Chapter XIV already at this point.

Examples for infinite-dimensional vector spaces occur in analysis. For instance, let V be the vector space of complex-valued C^∞ functions on $\mathbf{R}$. Let $D = d/dt$ be the derivative (if t is the variable). Then $D\colon V \to V$ is a linear map, and $\mathbf{C}[X]$ has the representation $\rho\colon \mathbf{C}[X] \to \mathrm{End}_{\mathbf{C}}(V)$ given by $P \mapsto P(D)$. A similar situation exists in several variables, when we let V be the vector space of C^∞ functions in n variables on an open set of $\mathbf{R}^n$. Then we let $D_i = \partial/\partial t_i$ be the partial derivative with respect to the i-th variable ($i = 1, \dots, n$). We obtain a representation

$$\rho\colon \mathbf{C}[X_1, \dots, X_n] \to \mathrm{End}_{\mathbf{C}}(V)$$

such that $\rho(X_i) = D_i$.

Example. Let H be a Hilbert space and let A be a bounded hermitian operator on A. Then one considers the homomorphism $\mathbf{R}[X] \to \mathbf{R}[A] \subset \mathrm{End}(H)$, from the polynomial ring into the algebra of endomorphisms of H, and one extends this homomorphism to the algebra of continuous functions on the spectrum of A. *Cf.* my *Real and Functional Analysis*, Springer Verlag, 1993.

Representations form a category as follows. We define a **morphism** of a representation $\rho\colon R \to \mathrm{End}_A(M)$ into a representation $\rho'\colon R \to \mathrm{End}_A(M')$, or in other words a **homomorphism of one representation of R to another**, to be an A-module homomorphism $h\colon M \to M'$ such that the following diagram is commutative for every $\alpha \in R$:

$$\begin{array}{ccc} M & \xrightarrow{h} & M' \\ {\scriptstyle \rho(\alpha)}\downarrow & & \downarrow{\scriptstyle \rho'(\alpha)} \\ M & \xrightarrow[h]{} & M' \end{array}$$

In the case when h is an isomorphism, then we may replace the above diagram by the commutative diagram

$$\begin{array}{ccc} & \overset{\rho}{\nearrow} & \mathrm{End}_A(M) \\ R & & \downarrow{\scriptstyle [h]} \\ & \underset{\rho'}{\searrow} & \mathrm{End}_A(M') \end{array}$$

where the symbol $[h]$ denotes conjugation by h, i.e. for $f \in \mathrm{End}_A(M)$ we have $[h]f = h \circ f \circ h^{-1}$.

Representations: from a monoid to the monoid algebra. Let G be a monoid. By a **representation of** G on an A-module M, we mean a homomorphism $\rho: G \to \mathrm{End}_A(M)$ of G into the multiplicative monoid of $\mathrm{End}_A(M)$. Then we may extend ρ to a homomorphism of the monoid algebra

$$A[G] \to \mathrm{End}_A(M),$$

by letting

$$\rho\left(\sum_{x \in G} a_x x\right) = \sum_{x \in G} a_x \rho(x).$$

It is immediately verified that this extension of ρ to $A[G]$ is a ring homomorphism, coinciding with the given ρ on elements of G.

Examples: modules over a group ring. The next examples will follow a certain pattern associated with groups of automorphisms. Quite generally, suppose we have some category of objects, and to each object K there is associated an abelian group $F(K)$, functorially with respect to isomorphisms. This means that if $\sigma: K \to K'$ is an isomorphism, then there is an associated isomorphism $F(\sigma): F(K') \to F(K')$ such that $F(\mathrm{id}) = \mathrm{id}$ and $F(\sigma\tau) = F(\sigma) \circ F(\tau)$. Then the group of automorphisms $\mathrm{Aut}(K)$ of an object operates on $F(K)$; that is, we have a natural homomorphism

$$\mathrm{Aut}(K) \to \mathrm{Aut}(F(K)) \quad \text{given by} \quad \sigma \mapsto F(\sigma).$$

Let $G = \mathrm{Aut}(K)$. Then $F(K)$ (written additively) can be made into a module over the group ring $\mathbf{Z}[G]$ as above. Given an element $\alpha = \sum a_\sigma \sigma \in \mathbf{Z}[G]$, with $a_\sigma \in \mathbf{Z}$, and an element $x \in F(K)$, we define

$$\alpha x = \sum a_\sigma F(\sigma)x.$$

The conditions defining a module are trivially satisfied. We list several concrete cases from mathematics at large, so there are no holds barred on the terminology.

Let K be a number field (i.e. a finite extension of the rational numbers). Let G be its group of automorphisms. Associated with K we have the following objects:

the ring of algebraic integers $\mathfrak{o}_K$;

the group of units $\mathfrak{o}_K^*$;

the group of ideal classes $C(K)$;

the group of roots of unity $\boldsymbol{\mu}(K)$.

Then G operates on each of those objects, and one problem is to determine the structure of these objects as $\mathbf{Z}[G]$-modules. Already for cyclotomic fields this

determination gives rise to substantial theories and to a number of unsolved problems.

Suppose that K is a Galois extension of k with Galois group G (see Chapter VI). Then we may view K itself as a module over the group ring $k[G]$. In Chapter VI, §13 we shall prove that K is isomorphic to $k[G]$ as module over $k[G]$ itself.

In topology, one considers a space X_0 and a finite covering X. Then $\mathrm{Aut}(X/X_0)$ operates on the homology of X, so this homology is a module over the group ring.

With more structure, suppose that X is a projective non-singular variety, say over the complex numbers. Then to X we can associate:

the group of divisor classes (Picard group) $\mathrm{Pic}(X)$;

in a given dimension, the group of cycle classes or Chow group $\mathrm{CH}^p(X)$;

the ordinary homology of X;

the sheaf cohomology in general.

If X is defined over a field K finitely generated over the rationals, we can associate a fancier cohomology defined algebraically by Grothendieck, and functorial with respect to the operation of Galois groups.

Then again all these objects can be viewed as modules over the group ring of automorphism groups, and major problems of mathematics consist in determining their structure. I direct the reader here to two surveys, which contain extensive bibliographies.

[CCFT 91] P. CASSOU-NOGUES, T. CHINBURG, A. FRÖHLICH, M. J. TAYLOR, L-functions and Galois modules, in *L-functions and Arithmetic* J. Coates and M. J. Taylor (eds.), *Proceedings of the Durham Symposium* July 1989, *London Math, Soc. Lecture Note Series* **153**, Cambridge University Press (1991), pp. 75-139

[La 82] S. LANG, Units and class groups in number theory and algebraic geometry, *Bull. AMS* **Vol. 6 No. 3** (1982), pp. 253-316

§3. DIRECT PRODUCTS AND SUMS OF MODULES

Let A be a ring. Let $\{M_i\}_{i \in I}$ be a family of modules. We defined their direct product as abelian groups in Chapter I, §9. Given an element $(x_i)_{i \in I}$ of the direct product, and $a \in A$, we define $a(x_i) = (ax_i)$. In other words, we multiply by an element a componentwise. Then the direct product $\prod M_i$ is an A-module. The reader will verify at once that it is also a **direct product** in the category of A-modules.

Similarly, let

$$M = \bigoplus_{i \in I} M_i$$

be their direct sum as abelian groups. We define on M a structure of A-module: If $(x_i)_{i \in I}$ is an element of M, i.e. a family of elements $x_i \in M_i$ such that $x_i = 0$ for almost all i, and if $a \in A$, then we define

$$a(x_i)_{i \in I} = (ax_i)_{i \in I},$$

that is we define multiplication by a componentwise. It is trivially verified that this is an operation of A on M which makes M into an A-module. If one refers back to the proof given for the existence of direct sums in the category of abelian groups, one sees immediately that this proof now extends in the same way to show that M is a direct sum of the family $\{M_i\}_{i \in I}$ as A-modules. (For instance, the map

$$\lambda_j : M_j \to M$$

such that $\lambda_j(x)$ has j-th component equal to x and i-th component equal to 0 for $i \neq j$ is now seen to be an A-homomorphism.)

This direct sum is a **coproduct in the category of A-modules**. Indeed, the reader can verify at once that given a family of A-homomorphisms $\{f_i : M_i \to N\}$, the map f defined as in the proof for abelian groups is also an A-isomorphism and has the required properties. See Proposition 7.1 of Chapter I.

When I is a finite set, there is a useful criterion for a module to be a direct product.

Proposition 3.1. *Let M be an A-module and n an integer $\geqq 1$. For each $i = 1, \ldots, n$ let $\varphi_i : M \to M$ be an A-homomorphism such that*

$$\sum_{i=1}^{n} \varphi_i = \text{id} \quad \textit{and} \quad \varphi_i \circ \varphi_j = 0 \qquad \textit{if } i \neq j.$$

Then $\varphi_i^2 = \varphi_i$ for all i. Let $M_i = \varphi_i(M)$, and let $\varphi : M \to \prod M_i$ be such that

$$\varphi(x) = (\varphi_1(x), \ldots, \varphi_n(x)).$$

Then φ is an A-isomorphism of M onto the direct product $\prod M_i$.

Proof. For each j, we have

$$\varphi_j = \varphi_j \circ \text{id} = \varphi_j \circ \sum_{i=1}^{n} \varphi_i = \varphi_j \circ \varphi_j = \varphi_j^2,$$

thereby proving the first assertion. It is clear that φ is an A-homomorphism. Let x be in its kernel. Since

$$x = \text{id}(x) = \sum_{i=1}^{n} \varphi_i(x)$$

we conclude that $x = 0$, so φ is injective. Given elements $y_i \in M_i$ for each $i = 1, \ldots, n$, let $x = y_1 + \cdots + y_n$. We obviously have $\varphi_j(y_i) = 0$ if $i \neq j$. Hence

$$\varphi_j(x) = y_j$$

for each $j = 1, \ldots, n$. This proves that φ is surjective, and concludes the proof of our proposition.

We observe that when I is a finite set, the direct sum and the direct product are equal.

Just as with abelian groups, we use the symbol $\oplus$ to denote direct sum.

Let M be a module over a ring A and let S be a subset of M. By a **linear combination** of elements of S (with coefficients in A) one means a sum

$$\sum_{x \in S} a_x x$$

where $\{a_x\}$ is a set of elements of A, almost all of which are equal to 0. These elements a_x are called the **coefficients** of the linear combination. Let N be the set of all linear combinations of elements of S. Then N is a submodule of M, for if

$$\sum_{x \in S} a_x x \quad \text{and} \quad \sum_{x \in S} b_x x$$

are two linear combinations, then their sum is equal to

$$\sum_{x \in S} (a_x + b_x)x,$$

and if $c \in A$, then

$$c\left(\sum_{x \in S} a_x x\right) = \sum_{x \in S} c a_x x,$$

and these elements are again linear combinations of elements of S. We shall call N the submodule **generated** by S, and we call S a set of **generators** for N. We sometimes write $N = A\langle S \rangle$. If S consists of one element x, the module generated by x is also written Ax, or simply (x), and sometimes we say that (x) is a **principal module**.

A module M is said to be **finitely generated**, or of **finite type**, or **finite** over A, if it has a finite number of generators.

A *subset* S of a module M is said to be **linearly independent** (over A) if whenever we have a linear combination

$$\sum_{x \in S} a_x x$$

which is equal to 0, then $a_x = 0$ for all $x \in S$. If S is linearly independent and if two linear combinations

$$\sum a_x x \quad \text{and} \quad \sum b_x x$$

are equal, then $a_x = b_x$ for all $x \in S$. Indeed, subtracting one from the other yields $\sum (a_x - b_x)x = 0$, whence $a_x - b_x = 0$ for all x. If S is linearly independent we shall also say that its elements are linearly independent. Similarly, a *family* $\{x_i\}_{i \in I}$ of elements of M is said to be linearly independent if whenever we have a linear combination

$$\sum_{i \in I} a_i x_i = 0,$$

then $a_i = 0$ for all i. A subset S (resp. a family $\{x_i\}$) is called **linearly dependent** if it is not linearly independent, i.e. if there exists a relation

$$\sum_{x \in S} a_x x = 0 \quad \text{resp.} \quad \sum_{i \in I} a_i x_i = 0$$

with not all a_x (resp. a_i) $= 0$. **Warning.** Let x be a single element of M which is linearly independent. Then the family $\{x_i\}_{i=1,\ldots,n}$ such that $x_i = x$ for all i is linearly dependent if $n > 1$, but the set consisting of x itself is linearly independent.

Let M be an A-module, and let $\{M_i\}_{i \in I}$ be a family of submodules. Since we have inclusion-homomorphisms

$$\lambda_i : M_i \to M$$

we have an induced homomorphism

$$\lambda_* : \bigoplus M_i \to M$$

which is such that for any family of elements $(x_i)_{i \in I}$, all but a finite number of which are 0, we have

$$\lambda_*((x_i)) = \sum_{i \in I} x_i.$$

If λ_* is an isomorphism, then we say that the family $\{M_i\}_{i \in I}$ is a **direct sum decomposition** of M. This is obviously equivalent to saying that every element of M has a unique expression as a sum

$$\sum x_i$$

with $x_i \in M_i$, and almost all $x_i = 0$. By abuse of notation, we also write

$$M = \bigoplus M_i$$

in this case.

If the family $\{M_i\}$ is such that every element of M has *some* expression as a sum $\sum x_i$ (not necessarily unique), then we write $M = \sum M_i$. In any case, if $\{M_i\}$ is an arbitrary family of submodules, the image of the homomorphism λ_* above is a submodule of M, which will be denoted by $\sum M_i$.

If M is a module and N, N' are two submodules such that $N + N' = M$ and $N \cap N' = 0$, then we have a module-isomorphism

$$M \approx N \oplus N',$$

just as with abelian groups, and similarly with a finite number of submodules.

We note, of course, that our discussion of abelian groups is a special case of our discussion of modules, simply by viewing abelian groups as modules over $\mathbf{Z}$. However, it seems usually desirable (albeit inefficient) to develop first some statements for abelian groups, and then point out that they are valid (obviously) for modules in general.

Let M, M', N be modules. Then we have an isomorphism of abelian groups

$$\boxed{\operatorname{Hom}_A(M \oplus M', N) \overset{\approx}{\leftrightarrow} \operatorname{Hom}_A(M, N) \times \operatorname{Hom}_A(M', N),}$$

and similarly

$$\boxed{\operatorname{Hom}_A(N, M \times M') \overset{\approx}{\leftrightarrow} \operatorname{Hom}_A(N, M) \times \operatorname{Hom}_A(N, M').}$$

The first one is obtained as follows. If $f : M \oplus M' \to N$ is a homomorphism, then f induces a homomorphism $f_1 : M \to N$ and a homomorphism $f_2 : M' \to N$ by composing f with the injections of M and M' into their direct sum respectively:

$$M \to M \oplus \{0\} \subset M \oplus M' \xrightarrow{f} N,$$

$$M' \to \{0\} \oplus M' \subset M \oplus M' \xrightarrow{f} N.$$

We leave it to the reader to verify that the association

$$f \mapsto (f_1, f_2)$$

gives an isomorphism as in the first box. The isomorphism in the second box is obtained in a similar way. Given homomorphisms

$$f_1 : N \to M$$

and

$$f_2 : N \to M'$$

we have a homomorphism $f: N \to M \times M'$ defined by

$$f(x) = (f_1(x), f_2(x)).$$

It is trivial to verify that the association

$$(f_1, f_2) \mapsto f$$

gives an isomorphism as in the second box.

Of course, the direct sum and direct product of two modules are isomorphic, but we distinguished them in the notation for the sake of functoriality, and to fit the infinite case, see Exercise 22.

Proposition 3.2. *Let* $0 \to M' \xrightarrow{f} M \xrightarrow{g} M'' \to 0$ *be an exact sequence of modules. The following conditions are equivalent:*

1. *There exists a homomorphism* $\varphi : M'' \to M$ *such that* $g \circ \varphi = \text{id}$.
2. *There exists a homomorphism* $\psi : M \to M'$ *such that* $\psi \circ f = \text{id}$.

If these conditions are satisfied, then we have isomorphisms:

$$M = \operatorname{Im} f \oplus \operatorname{Ker} \psi, \qquad M = \operatorname{Ker} g \oplus \operatorname{Im} \varphi,$$

$$M \approx M' \oplus M''.$$

Proof. Let us write the homomorphisms on the right:

$$M \underset{\varphi}{\overset{g}{\rightleftarrows}} M'' \to 0.$$

Let $x \in M$. Then

$$x - \varphi(g(x))$$

is in the kernel of g, and hence $M = \operatorname{Ker} g + \operatorname{Im} \varphi$.

This sum is direct, for if

$$x = y + z$$

with $y \in \operatorname{Ker} g$ and $z \in \operatorname{Im} \varphi$, $z = \varphi(w)$ with $w \in M''$, and applying g yields $g(x) = w$. Thus w is uniquely determined by x, and therefore z is uniquely determined by x. Hence so is y, thereby proving the sum is direct.

The arguments concerning the other side of the sequence are similar and will be left as exercises, as well as the equivalence between our conditions. When these conditions are satisfied, the exact sequence of Proposition 3.2 is said to **split**. One also says that ψ **splits** f and φ **splits** g.

Abelian categories

Much in the theory of modules over a ring is arrow-theoretic. In fact, one needs only the notion of kernel and cokernel (factor modules). One can axiomatize the special notion of a category in which many of the arguments are valid, especially the arguments used in this chapter. Thus we give this axiomatization now, although for concreteness, at the beginning of the chapter, we continue to use the language of modules. Readers should strike their own balance when they want to slide into the more general framework.

Consider first a category $\mathfrak{a}$ such that $\mathrm{Mor}(E, F)$ is an abelian group for each pair of objects E, F of $\mathfrak{a}$, satisfying the following two conditions:

AB 1. The law of composition of morphisms is bilinear, and there exists a zero object 0, i.e. such that $\mathrm{Mor}(0, E)$ and $\mathrm{Mor}(E, 0)$ have precisely one element for each object E.

AB 2. Finite products and finite coproducts exist in the category.

Then we say that $\mathfrak{a}$ is an **additive category**.

Given a morphism $E \xrightarrow{f} F$ in $\mathfrak{a}$, we define a **kernel** of f to be a morphism $E' \to E$ such that for all objects X in the category, the following sequence is exact:

$$0 \to \mathrm{Mor}(X, E') \to \mathrm{Mor}(X, E) \to \mathrm{Mor}(X, F).$$

We define a **cokernel** for f to be a morphism $F \to F''$ such that for all objects X in the category, the following sequence is exact:

$$0 \to \mathrm{Mor}(F'', X) \to \mathrm{Mor}(F, X) \to \mathrm{Mor}(E, X).$$

It is immediately verified that kernels and cokernels are universal in a suitable category, and hence uniquely determined up to a unique isomorphism if they exist.

AB 3. Kernels and cokernels exist.

AB 4. If $f: E \to F$ is a morphism whose kernel is 0, then f is the kernel of its cokernel. If $f: E \to F$ is a morphism whose cokernel is 0, then f is the cokernel of its kernel. A morphism whose kernel and cokernel are 0 is an isomorphism.

A category $\mathfrak{a}$ satisfying the above four axioms is called an **abelian category**.

In an abelian caegory, the group of morphisms is usually denoted by Hom, so for two objects E, F we write

$$\mathrm{Mor}(E, F) = \mathrm{Hom}(E, F).$$

The morphisms are usually called **homomorphisms**. Given an exact sequence

$$0 \to M' \to M,$$

we say that M' is a **subobject** of M, or that the homomorphism of M' into M is a **monomorphism**. Dually, in an exact sequence

$$M \to M'' \to 0,$$

we say that M'' is a **quotient object** of M, or that the homomorphism of M to M'' is an **epimorphism**, instead of saying that it is surjective as in the category of modules. Although it is convenient to think of modules and abelian groups to construct proofs, usually such proofs will involve only arrow-theoretic arguments, and will therefore apply to any abelian category. However, all the abelian categories we shall meet in this book will have elements, and the kernels and cokernels will be defined in a natural fashion, close to those for modules, so readers may restrict their attention to these concrete cases.

Examples of abelian categories. Of course, modules over a ring form an abelian category, the most common one. Finitely generated modules over a Noetherian ring form an abelian category, to be studied in Chapter X.

Let k be a field. We consider pairs (V, A) consisting of a finite-dimensional vector space V over k, and an endomorphism $A: V \to V$. By a **homomorphism (morphism)** of such pairs $f: (V, A) \to (W, B)$ we mean a k-homomorphism $f: V \to W$ such that the following diagram is commutative:

$$\begin{array}{ccc} V & \xrightarrow{f} & W \\ {\scriptstyle A}\downarrow & & \downarrow{\scriptstyle B} \\ V & \xrightarrow[f]{} & W \end{array}$$

It is routinely verified that such pairs and the above defined morphisms form an abelian category. Its elements will be studied in Chapter XIV.

Let k be a field and let G be a group. Let $\text{Mod}_k(G)$ be the category of finite-dimensional vector spaces V over k, with an operation of G on V, i.e. a homomorphism $G \to \text{Aut}_k(V)$. A homomorphism (morphism) in that category is a k-homomorphism $f: V \to W$ such that $f(ax) = af(x)$ for all $x \in V$ and $a \in G$. It is immediate that $\text{Mod}_k(G)$ is an abelian category. This category will be studied especially in Chapter XVIII.

In Chapter XX, §1 we shall consider the category of complexes of modules over a ring. This category of complexes is an abelian category.

In topology and differential geometry, the category of vector bundles over a topological space is an abelian category.

Sheaves of abelian groups over a topological space form an abelian category, which will be defined in Chapter XX, §6.

§4. FREE MODULES

Let M be a module over a ring A and let S be a subset of M. We shall say that S is a **basis** of M if S is not empty, if S generates M, and if S is linearly independent. If S is a basis of M, then in particular $M \neq \{0\}$ if $A \neq \{0\}$ and every element of M has a unique expression as a linear combination of elements of S. Similarly, let $\{x_i\}_{i \in I}$ be a non-empty family of elements of M. We say that it is a **basis** of M if it is linearly independent and generates M.

If A is a ring, then as a module over itself, A admits a basis, consisting of the unit element 1.

Let I be a non-empty set, and for each $i \in I$, let $A_i = A$, viewed as an A-module. Let

$$F = \bigoplus_{i \in I} A_i.$$

Then F admits a basis, which consists of the elements e_i of F whose i-th component is the unit element of A_i, and having all other components equal to 0.

By a **free** module we shall mean a module which admits a basis, or the zero module.

Theorem 4.1. *Let A be a ring and M a module over A. Let I be a non-empty set, and let $\{x_i\}_{i \in I}$ be a basis of M. Let N be an A-module, and let $\{y_i\}_{i \in I}$ be a family of elements of N. Then there exists a unique homomorphism $f : M \to N$ such that $f(x_i) = y_i$ for all i.*

Proof. Let x be an element of M. There exists a unique family $\{a_i\}_{i \in I}$ of elements of A such that

$$x = \sum_{i \in I} a_i x_i.$$

We define

$$f(x) = \sum a_i y_i.$$

It is then clear that f is a homomorphism satisfying our requirements, and that it is the unique such, because we must have

$$f(x) = \sum a_i f(x_i).$$

Corollary 4.2. *Let the notation be as in the theorem, and assume that $\{y_i\}_{i \in I}$ is a basis of N. Then the homomorphism f is an isomorphism, i.e. a module-isomorphism.*

Proof. By symmetry, there exists a unique homomorphism

$$g : N \to M$$

such that $g(y_i) = x_i$ for all i, and $f \circ g$ and $g \circ f$ are the respective identity mappings.

Corollary 4.3. *Two modules having bases whose cardinalities are equal are isomorphic.*

Proof. Clear.

We shall leave the proofs of the following statements as exercises.

Let M be a free module over A, with basis $\{x_i\}_{i \in I}$, so that

$$M = \bigoplus_{i \in I} Ax_i.$$

Let $\mathfrak{a}$ be a two sided ideal of A. Then $\mathfrak{a}M$ is a submodule of M. Each $\mathfrak{a}x_i$ is a submodule of Ax_i. *We have an isomorphism* (of A-modules)

$$\boxed{M/\mathfrak{a}M \approx \bigoplus_{i \in I} Ax_i/\mathfrak{a}x_i.}$$

Furthermore, each $Ax_i/\mathfrak{a}x_i$ is isomorphic to $A/\mathfrak{a}$, as A-module.

Suppose in addition that A is commutative. Then $A/\mathfrak{a}$ is a ring. Furthermore $M/\mathfrak{a}M$ is a free module over $A/\mathfrak{a}$, and each $Ax_i/\mathfrak{a}x_i$ is free over $A/\mathfrak{a}$. If $\bar{x}_i$ is the image of x_i under the canonical homomorphism

$$Ax_i \to Ax_i/\mathfrak{a}x_i,$$

then the single element $\bar{x}_i$ is a basis of $Ax_i/\mathfrak{a}x_i$ over $A/\mathfrak{a}$.

All of these statements should be easily verified by the reader. Now let A be an arbitrary commutative ring. A module M is called **principal** if there exists an element $x \in M$ such that $M = Ax$. The map

$$a \mapsto ax \text{ (for } a \in A)$$

is an A-module homomorphism of A onto M, whose kernel is a left ideal $\mathfrak{a}$, and inducing an isomorphism of A-modules

$$A/\mathfrak{a} \approx M.$$

Let M be a finitely generated module, with generators $\{v_1, \ldots, v_n\}$. Let F be a free module with basis $\{e_1, \ldots, e_n\}$. Then there is a unique surjective homomorphism $f: F \to M$ such that $f(e_i) = v_i$. The kernel of f is a submodule M_1. Under certain conditions, M_1 is finitely generated (*cf.* Chapter X, §1 on Noetherian rings), and the process can be continued. The systematic study of this process will be carried out in the chapters on resolutions of modules and homology.

Of course, even if M is not finitely generated, one can carry out a similar construction, by using an arbitrary indexing set. Indeed, let $\{v_i\}$ $(i \in I)$ be a family of generators. For each i, let F_i be free with basis consisting of a single element e_i, so $F_i \approx A$. Let F be the direct sum of the modules F_i $(i \in I)$, as in Proposition 3.1. Then we obtain a surjective homomorphism $f: F \to M$ such that $f(e_i) = v_i$. Thus every module is a factor module of a free module.

Just as we did for abelian groups in Chapter I, §7, we can also define the **free module** over a ring A **generated by a non-empty set** S. We let $A\langle S\rangle$ be the set of functions $\varphi : S \to A$ such that $\varphi(x) = 0$ for almost all $x \in S$. If $a \in A$ and $x \in S$, we denote by ax the map φ such that $\varphi(x) = a$ and $\varphi(y) = 0$ for $y \neq x$. Then as for abelian groups, given $\varphi \in A\langle S\rangle$ there exist elements $a_i \in A$ and $x_i \in S$ such that

$$\varphi = a_1x_1 + \cdots + a_nx_n.$$

It is immediately verified that the family of functions $\{\delta_x\}$ $(x \in S)$ such that $\delta_x(x) = 1$ and $\delta_x(y) = 0$ for $y \neq x$ form a basis for $A\langle S\rangle$. In other words, the expression of φ as $\sum a_ix_i$ above is unique. This construction can be applied when S is a group or a monoid G, and gives rise to the group algebra as in Chapter II, §5.

Projective modules

There exists another important type of module closely related to free modules, which we now discuss.

Let A be a ring and P a module. The following properties are equivalent, and define what it means for P to be a **projective module**.

P 1. Given a homomorphism $f: P \to M''$ and surjective homomorphism $g: M \to M''$, there exists a homomorphism $h: P \to M$ making the following diagram commutative.

$$\begin{array}{ccccc} & & P & & \\ & \swarrow^{h} & \downarrow{\scriptstyle f} & & \\ M & \xrightarrow[g]{} & M'' & \longrightarrow & 0 \end{array}$$

P 2. Every exact sequence $0 \to M' \to M'' \to P \to 0$ splits.

P 3. There exists a module M such that $P \oplus M$ is free, or in words, P is a direct summand of a free module.

P 4. The functor $M \mapsto \operatorname{Hom}_A(P, M)$ is exact.

We prove the equivalence of the four conditions.

Assume **P 1**. Given the exact sequence of **P 2**, we consider the map $f = \text{id}$ in the diagram

$$\begin{array}{ccccc} & & P & & \\ & \swarrow^{h} & \downarrow \text{id} & & \\ M'' & \longrightarrow & P & \longrightarrow & 0 \end{array}$$

Then h gives the desired splitting of the sequence.

Assume **P 2**. Then represent P as a quotient of a free module (cf. Exercise 1) $F \to P \to 0$, and apply **P 2** to this sequence to get the desired splitting, which represents F as a direct sum of P and some module.

Assume **P 3**. Since $\text{Hom}_A(X \oplus Y, M) = \text{Hom}_A(X, M) \oplus \text{Hom}_A(Y, M)$, and since $M \mapsto \text{Hom}_A(F, M)$ is an exact functor if F is free, it follows that $\text{Hom}_A(P, M)$ is exact when P is a direct summand of a free module, which proves **P 4**.

Assume **P 4**. The proof of **P 1** will be left as an exercise.

Examples. It will be proved in the next section that a vector space over a field is always free, i.e. has a basis. Under certain circumstances, it is a theorem that projective modules are free. In §7 we shall prove that a finitely generated projective module over a principal ring is free. In Chapter X, Theorem 4.4 we shall prove that such a module over a local ring is free; in Chapter XVI, Theorem 3.8 we shall prove that a finite flat module over a local ring is free; and in Chapter XXI, Theorem 3.7, we shall prove the Quillen-Suslin theorem that if $A = k[X_1, \dots, X_n]$ is the polynomial ring over a field k, then every finite projective module over A is free.

Projective modules give rise to the Grothendieck group. Let A be a ring. Isomorphism classes of finite projective modules form a monoid. Indeed, if P is finite projective, let $[P]$ denote its isomorphism class. We define

$$[P] + [Q] = [P \oplus Q].$$

This sum is independent of the choice of representatives P, Q in their class. The conditions defining a monoid are immediately verified. The corresponding Grothendieck group is denoted by $K(A)$.

We can impose a further equivalence relation that P is equivalent to P' if there exist finite free modules F and F' such that $P \oplus F$ is isomorphic to $P' \oplus F'$. Under this equivalence relation we obtain another group denoted by $K_0(A)$. If A is a Dedekind ring (Chapter II, §1 and Exercises 13–19) it can be shown that this group is isomorphic in a natural way with the group of ideal classes $\text{Pic}(A)$ (defined in Chapter II, §1). See Exercises 11, 12, 13. It is also a

problem to determine $K_0(A)$ for as many rings as possible, as explicitly as possible. Algebraic number theory is concerned with $K_0(A)$ when A is the ring of algebraic integers of a number field. The Quillen-Suslin theorem shows if A is the polynomial ring as above, then $K_0(A)$ is trivial.

Of course one can carry out a similar construction with all finite modules. Let $[M]$ denote the isomorphism class of a finite module M. We define the sum to be the direct sum. Then the isomorphism classes of modules over the ring form a monoid, and we can associate to this monoid its Grothendieck group. This construction is applied especially when the ring is commutative. There are many variations on this theme. See for instance the book by Bass, *Algebraic K-theory*, Benjamin, 1968.

There is a variation of the definition of Grothendieck group as follows. Let F be the free abelian group generated by isomorphism classes of finite modules over a ring R, or of modules of bounded cardinality so that we deal with sets. In this free abelian group we let Γ be the subgroup generated by all elements

$$[M] - [M'] - [M'']$$

for which there exists an exact sequence $0 \to M' \to M \to M'' \to 0$. The factor group F/Γ is called the **Grothendieck group** $K(R)$. We shall meet this group again in §8, and in Chapter XX, §3. Note that we may form a similar Grothendieck group with any family of modules such that M is in the family if and only if M' and M'' are in the family. Taking for the family finite projective modules, one sees easily that the two possible definitions of the Grothendieck group coincide in that case.

§5. VECTOR SPACES

A module over a field is called a **vector space**.

Theorem 5.1. *Let V be a vector space over a field K, and assume that $V \neq \{0\}$. Let Γ be a set of generators of V over K and let S be a subset of Γ which is linearly independent. Then there exists a basis $\mathfrak{B}$ of V such that $S \subset \mathfrak{B} \subset \Gamma$.*

Proof. Let $\mathfrak{T}$ be the set whose elements are subsets T of Γ which contain S and are linearly independent. Then $\mathfrak{T}$ is not empty (it contains S), and we contend that $\mathfrak{T}$ is inductively ordered. Indeed, if $\{T_i\}$ is a totally ordered subset

of $\mathfrak{T}$ (by ascending inclusion), then $\bigcup T_i$ is again linearly independent and contains S. By Zorn's lemma, let $\mathfrak{B}$ be a maximal element of $\mathfrak{T}$. Then $\mathfrak{B}$ is linearly independent. Let W be the subspace of V generated by $\mathfrak{B}$. If $W \neq V$, there exists some element $x \in \Gamma$ such that $x \notin W$. Then $\mathfrak{B} \cup \{x\}$ is linearly independent, for given a linear combination

$$\sum_{y \in \mathfrak{B}} a_y y + bx = 0, \qquad a_y, b \in K,$$

we must have $b = 0$, otherwise we get

$$x = -\sum_{y \in \mathfrak{B}} b^{-1} a_y y \in W.$$

By construction, we now see that $a_y = 0$ for all $y \in \mathfrak{B}$, thereby proving that $\mathfrak{B} \cup \{x\}$ is linearly independent, and contradicting the maximality of $\mathfrak{B}$. It follows that $W = V$, and furthermore that $\mathfrak{B}$ is not empty since $V \neq \{0\}$. This proves our theorem.

If V is a vector space $\neq \{0\}$, then in particular, we see that every set of linearly independent elements of V can be extended to a basis, and that a basis may be selected from a given set of generators.

Theorem 5.2. *Let V be a vector space over a field K. Then two bases of V over K have the same cardinality.*

Proof. Let us first assume that there exists a basis of V with a finite number of elements, say $\{v_1, \dots, v_m\}$, $m \geqq 1$. We shall prove that any other basis must also have m elements. For this it will suffice to prove: If $w_1, \dots, w_n$ are elements of V which are linearly independent over K, then $n \leqq m$ (for we can then use symmetry). We proceed by induction. There exist elements $c_1, \dots, c_m$ of K such that

$$w_1 = c_1 v_1 + \cdots + c_m v_m, \tag{1}$$

and some c_i, say c_1, is not equal to 0. Then v_1 lies in the space generated by $w_1, v_2, \dots, v_m$ over K, and this space must therefore be equal to V itself. Furthermore, $w_1, v_2, \dots, v_m$ are linearly independent, for suppose $b_1, \dots, b_m$ are elements of K such that

$$b_1 w_1 + b_2 v_2 + \cdots + b_m v_m = 0.$$

If $b_1 \neq 0$, divide by b_1 and express w_1 as a linear combination of $v_2, \dots, v_m$. Subtracting from (1) would yield a relation of linear dependence among the v_i, which is impossible. Hence $b_1 = 0$, and again we must have all $b_i = 0$ because the v_i are linearly independent.

Suppose inductively that after a suitable renumbering of the v_i, we have found $w_1, \dots, w_r$ $(r < n)$ such that

$$\{w_1, \dots, w_r, v_{r+1}, \dots, v_m\}$$

is a basis of V. We express w_{r+1} as a linear combination

$$w_{r+1} = c_1 w_1 + \cdots + c_r w_r + c_{r+1} v_{r+1} + \cdots + c_m v_m \tag{2}$$

with $c_i \in K$. The coefficients of the v_i in this relation cannot all be 0; otherwise there would be a linear dependence among the w_j. Say $c_{r+1} \neq 0$. Using an argument similar to that used above, we can replace v_{r+1} by w_{r+1} and still have a basis of V. This means that we can repeat the procedure until $r = n$, and therefore that $n \leqq m$, thereby proving our theorem.

We shall leave the general case of an infinite basis as an exercise to the reader. [*Hint*: Use the fact that a finite number of elements in one basis is contained in the space generated by a finite number of elements in another basis.]

If a vector space V admits one basis with a finite number of elements, say m, then we shall say that V is **finite dimensional** and that m is its **dimension**. In view of Theorem 5.2, we see that m is the number of elements in *any* basis of V. If $V = \{0\}$, then we define its dimension to be 0, and say that V is 0-dimensional. We abbreviate "dimension" by "dim" or "$\dim_K$" if the reference to K is needed for clarity.

When dealing with vector spaces over a field, we use the words **subspace** and **factor space** instead of **submodule** and **factor module**.

Theorem 5.3. *Let V be a vector space over a field K, and let W be a subspace. Then*

$$\dim_K V = \dim_K W + \dim_K V/W.$$

If $f: V \to U$ is a homomorphism of vector spaces over K, then

$$\dim V = \dim \operatorname{Ker} f + \dim \operatorname{Im} f.$$

Proof. The first statement is a special case of the second, taking for f the canonical map. Let $\{u_i\}_{i \in I}$ be a basis of $\operatorname{Im} f$, and let $\{w_j\}_{j \in J}$ be a basis of $\operatorname{Ker} f$. Let $\{v_i\}_{i \in I}$ be a family of elements of V such that $f(v_i) = u_i$ for each $i \in I$. We contend that

$$\{v_i, w_j\}_{i \in I, j \in J}$$

is a basis for V. This will obviously prove our assertion.

Let x be an element of V. Then there exist elements $\{a_i\}_{i \in I}$ of K almost all of which are 0 such that

$$f(x) = \sum_{i \in I} a_i u_i.$$

Hence $f(x - \sum a_i v_i) = f(x) - \sum a_i f(v_i) = 0$. Thus

$$x - \sum a_i v_i$$

is in the kernel of f, and there exist elements $\{b_j\}_{j \in J}$ of K almost all of which are 0 such that

$$x - \sum a_i v_i = \sum b_j w_j.$$

From this we see that $x = \sum a_i v_i + \sum b_j w_j$, and that $\{v_i, w_j\}$ generates V. It remains to be shown that the family $\{v_i, w_j\}$ is linearly independent. Suppose that there exist elements c_i, d_j such that

$$0 = \sum c_i v_i + \sum d_j w_j.$$

Applying f yields

$$0 = \sum c_i f(v_i) = \sum c_i u_i,$$

whence all $c_i = 0$. From this we conclude at once that all $d_j = 0$, and hence that our family $\{v_i, w_j\}$ is a basis for V over K, as was to be shown.

Corollary 5.4. *Let V be a vector space and W a subspace. Then*

$$\dim W \leqq \dim V.$$

If V is finite dimensional and $\dim W = \dim V$ *then $W = V$.*

Proof. Clear.

§6. THE DUAL SPACE AND DUAL MODULE

Let E be a free module over a commutative ring A. We view A as a free module of rank 1 over itself. By the **dual module** $E^\vee$ of E we shall mean the module $\text{Hom}(E, A)$. Its elements will be called **functionals**. Thus a functional on E is an A-linear map $f: E \to A$. If $x \in E$ and $f \in E^\vee$, we sometimes denote $f(x)$ by $\langle x, f \rangle$. Keeping x fixed, we see that the symbol $\langle x, f \rangle$ as a function of $f \in E^\vee$ is A-linear in its second argument, and hence that x induces a linear map on $E^\vee$, which is 0 if and only if $x = 0$. Hence we get an injection $E \to E^{\vee\vee}$ which is not always a surjection.

Let $\{x_i\}_{i \in I}$ be a basis of E. For each $i \in I$ let f_i be the unique functional such that $f_i(x_j) = \delta_{ij}$ (in other words, 1 if $i = j$ and 0 if $i \neq j$). Such a linear map exists by general properties of bases (Theorem 4.1).

Theorem 6.1. *Let E be a finite free module over the commutative ring A, of finite dimension n. Then $E^\vee$ is also free, and* $\dim E^\vee = n$. *If $\{x_1, \ldots, x_n\}$ is a basis for E, and f_i is the functional such that $f_i(x_j) = \delta_{ij}$, then $\{f_1, \ldots, f_n\}$ is a basis for $E^\vee$.*

Proof. Let $f \in E^\vee$ and let $a_i = f(x_i)$ $(i = 1, \ldots, n)$. We have

$$f(c_1x_1 + \cdots + c_nx_n) = c_1f(x_1) + \cdots + c_nf(x_n).$$

Hence $f = a_1f_1 + \cdots + a_nf_n$, and we see that the f_i generate $E^\vee$. Furthermore, they are linearly independent, for if

$$b_1f_1 + \cdots + b_nf_n = 0$$

with $b_i \in K$, then evaluating the left-hand side on x_i yields

$$b_if_i(x_i) = 0,$$

whence $b_i = 0$ for all i. This proves our theorem.

Given a basis $\{x_i\}$ $(i = 1, \ldots, n)$ as in the theorem, we call the basis $\{f_i\}$ the **dual basis**. In terms of these bases, we can express an element A of E with coordinates $(a_1, \ldots, a_n)$, and an element B of $E^\vee$ with coordinates $(b_1, \ldots, b_n)$, such that

$$A = a_1x_1 + \cdots + a_nx_n, \qquad B = b_1f_1 + \cdots + b_nf_n.$$

Then in terms of these coordinates, we see that

$$\langle A, B \rangle = a_1b_1 + \cdots + a_nb_n = A \cdot B$$

is the usual dot product of n-tuples.

Corollary 6.2. *When E is free finite dimensional, then the map $E \to E^{\vee\vee}$ which to each $x \in V$ associates the functional $f \mapsto \langle x, f \rangle$ on $E^\vee$ is an isomorphism of E onto $E^{\vee\vee}$.*

Proof. Note that since $\{f_1, \ldots, f_n\}$ is a basis for $E^\vee$, it follows from the definitions that $\{x_1, \ldots, x_n\}$ is the dual basis in E, so $E = E^{\vee\vee}$.

Theorem 6.3. *Let U, V, W be finite free modules over the commutative ring A, and let*

$$0 \to W \xrightarrow{\lambda} V \xrightarrow{\varphi} U \to 0$$

be an exact sequence of A-homomorphisms. Then the induced sequence

$$0 \to \operatorname{Hom}_A(U, A) \to \operatorname{Hom}_A(V, A) \to \operatorname{Hom}_A(W, A) \to 0$$

i.e.

$$0 \to U^\vee \to V^\vee \to W^\vee \to 0$$

is also exact.

Proof. This is a consequence of **P2**, because a free module is projective.

We now consider properties which have specifically to do with vector spaces, because we are going to take factor spaces. So we assume that we deal with vector spaces over a field K.

Let V, V' be two vector spaces, and suppose given a mapping

$$V \times V' \to K$$

denoted by

$$(x, x') \mapsto \langle x, x' \rangle$$

for $x \in V$ and $x' \in V'$. We call the mapping **bilinear** if for each $x \in V$ the function $x' \mapsto \langle x, x' \rangle$ is linear, and similarly for each $x' \in V'$ the function $x \mapsto \langle x, x' \rangle$ is linear. An element $x \in V$ is said to be **orthogonal** (or **perpendicular**) to a subset S' of V' if $\langle x, x' \rangle = 0$ for all $x' \in S'$. We make a similar definition in the opposite direction. It is clear that the set of $x \in V$ orthogonal to S' is a subspace of V.

We define the **kernel** of the bilinear map on the left to be the subspace of V which is orthogonal to V', and similarly for the kernel on the right.

Given a bilinear map as above,

$$V \times V' \to K,$$

let W' be its kernel on the right and let W be its kernel on the left. Let x' be an element of V'. Then x' gives rise to a functional on V, by the rule $x \mapsto \langle x, x' \rangle$, and this functional obviously depends only on the coset of x' modulo W'; in other words, if $x'_1 \equiv x'_2 \pmod{W'}$, then the functionals $x \mapsto \langle x, x'_1 \rangle$ and $x \mapsto \langle x, x'_2 \rangle$ are equal. Hence we get a homomorphism

$$V' \to V^\vee$$

whose kernel is precisely W' by definition, whence an injective homomorphism

$$0 \to V'/W' \to V^\vee.$$

Since all the functionals arising from elements of V' vanish on W, we can view them as functionals on V/W, i.e. as elements of $(V/W)^\vee$. So we actually get an injective homomorphism

$$0 \to V'/W' \to (V/W)^\vee.$$

One could give a name to the homomorphism

$$g : V' \to V^\vee$$

such that

$$\langle x, x' \rangle = \langle x, g(x') \rangle$$

for all $x \in V$ and $x' \in V'$. However, it will usually be possible to describe it by an arrow and call it the induced map, or the natural map. Giving a name to it would tend to make the terminology heavier than necessary.

Theorem 6.4. *Let $V \times V' \to K$ be a bilinear map, let W, W' be its kernels on the left and right respectively, and assume that V'/W' is finite dimensional. Then the induced homomorphism $V'/W' \to (V/W)^\vee$ is an isomorphism.*

Proof. By symmetry, we have an induced homomorphism

$$V/W \to (V'/W')^\vee$$

which is injective. Since

$$\dim(V'/W')^\vee = \dim V'/W'$$

it follows that V/W is finite dimensional. From the above injective homomorphism and the other, namely

$$0 \to V'/W' \to (V/W)^\vee,$$

we get the inequalities

$$\dim V/W \leqq \dim V'/W'$$

and

$$\dim V'/W' \leqq \dim V/W,$$

whence an equality of dimensions. Hence our homomorphisms are surjective and inverse to each other, thereby proving the theorem.

Remark 1. Theorem 6.4 is the analogue for vector spaces of the duality Theorem 9.2 of Chapter I.

Remark 2. Let A be a commutative ring and let E be an A-module. Then we may form two types of dual:

$E^\wedge = \mathrm{Hom}(E, \mathbf{Q}/\mathbf{Z})$, viewing E as an abelian group;

$E^\vee = \mathrm{Hom}_A(E, A)$, viewing E as an A-module.

Both are called **dual**, and they usually are applied in different contexts. For instance, $E^\vee$ will be considered in Chapter XIII, while $E^\wedge$ will be considered in the theory of injective modules, Chapter XX, §4. For an example of dual module $E^\vee$ see Exercise 11. If by any chance the two duals arise together and there is need to distinguish between them, then we may call $E^\wedge$ the **Pontrjagin dual**.

Indeed, in the theory of topological groups G, the group of continuous homomorphisms of G into $\mathbf{R}/\mathbf{Z}$ is the classical Pontrjagin dual, and is classically denoted by $G^\wedge$, so I find the preservation of that terminology appropriate.

Instead of $\mathbf{R}/\mathbf{Z}$ one may take other natural groups isomorphic to $\mathbf{R}/\mathbf{Z}$. The most common such group is the group of complex numbers of absolute value 1, which we denote by $\mathbf{S}^1$. The isomorphism with $\mathbf{R}/\mathbf{Z}$ is given by the map

$$x \mapsto e^{2\pi i x}.$$

Remark 3. A bilinear map $V \times V \to K$ for which $V' = V$ is called a **bilinear form**. We say that the form is **non-singular** if the corresponding maps

$$V' \to V^\vee \quad \text{and} \quad V \to (V')^\vee$$

are isomorphisms. Bilinear maps and bilinear forms will be studied at greater length in Chapter XV. See also Exercise 33 of Chapter XIII for a nice example.

§7. MODULES OVER PRINCIPAL RINGS

Throughout this section, we assume that R is a principal entire ring. All modules are over R, and homomorphisms are R-homomorphisms, unless otherwise specified.

The theorems will generalize those proved in Chapter I for abelian groups. We shall also point out how the proofs of Chapter I can be adjusted with substitutions of terminology so as to yield proofs in the present case.

Let F be a free module over R, with a basis $\{x_i\}_{i \in I}$. Then the cardinality of I is uniquely determined, and is called the **dimension** of F. We recall that this is proved, say by taking a prime element p in R, and observing that F/pF is a vector space over the field R/pR, whose dimension is precisely the cardinality of I. We may therefore speak of the dimension of a free module over R.

Theorem 7.1. *Let F be a free module, and M a submodule. Then M is free, and its dimension is less than or equal to the dimension of F.*

Proof. For simplicity, we give the proof when F has a finite basis $\{x_i\}$, $i = 1, \dots, n$. Let M_r be the intersection of M with $(x_1, \dots, x_r)$, the module generated by $x_1, \dots, x_r$. Then $M_1 = M \cap (x_1)$ is a submodule of (x_1), and is therefore of type (a_1x_1) with some $a_1 \in R$. Hence M_1 is either 0 or free, of dimension 1. Assume inductively that M_r is free of dimension $\leqq r$. Let $\mathfrak{a}$ be the set consisting of all elements $a \in R$ such that there exists an element $x \in M$ which can be written

$$x = b_1x_1 + \cdots + b_rx_r + ax_{r+1}$$

with $b_i \in R$. Then $\mathfrak{a}$ is obviously an ideal, and is principal, generated say by an element a_{r+1}. If $a_{r+1} = 0$, then $M_{r+1} = M_r$ and we are done with the inductive step. If $a_{r+1} \neq 0$, let $w \in M_{r+1}$ be such that the coefficient of w with respect to x_{r+1} is a_{r+1}. If $x \in M_{r+1}$ then the coefficient of x with respect to x_{r+1} is divisible by a_{r+1}, and hence there exists $c \in R$ such that $x - cw$ lies in M_r. Hence

$$M_{r+1} = M_r + (w).$$

On the other hand, it is clear that $M_r \cap (w)$ is 0, and hence that this sum is direct, thereby proving our theorem. (For the infinite case, see Appendix 2, §2.)

Corollary 7.2. *Let E be a finitely generated module and E' a submodule. Then E' is finitely generated.*

Proof. We can represent E as a factor module of a free module F with a finite number of generators: If $v_1, \ldots, v_n$ are generators of E, we take a free module F with basis $\{x_1, \ldots, x_n\}$ and map x_i on v_i. The inverse image of E' in F is a submodule, which is free, and finitely generated, by the theorem. Hence E' is finitely generated. The assertion also follows using simple properties of Noetherian rings and modules.

If one wants to translate the proofs of Chapter I, then one makes the following definitions. A free 1-dimensional module over R is called **infinite cyclic**. An infinite cyclic module is isomorphic to R, viewed as module over itself. Thus every non-zero submodule of an infinite cyclic module is infinite cyclic. The proof given in Chapter I for the analogue of Theorem 7.1 applies without further change.

Let E be a module. We say that E is a **torsion** module if given $x \in E$, there exists $a \in R$, $a \neq 0$, such that $ax = 0$. The generalization of **finite abelian group** is **finitely generated torsion module**. An element x of E is called a **torsion element** if there exists $a \in R$, $a \neq 0$, such that $ax = 0$.

Let E be a module. We denote by E_{tor} the submodule consisting of all torsion elements of E, and call it the **torsion submodule** of E. If $E_{\text{tor}} = 0$, we say that E is **torsion free**.

Theorem 7.3. *Let E be finitely generated. Then E/E_{tor} is free. There exists a free submodule F of E such that E is a direct sum*

$$E = E_{\text{tor}} \oplus F.$$

The dimension of such a submodule F is uniquely determined.

Proof. We first prove that E/E_{tor} is torsion free. If $x \in E$, let $\bar{x}$ denote its residue class mod E_{tor}. Let $b \in R$, $b \neq 0$ be such that $b\bar{x} = 0$. Then $bx \in E_{\text{tor}}$, and hence there exists $c \in R$, $c \neq 0$, such that $cbx = 0$. Hence $x \in E_{\text{tor}}$ and $\bar{x} = 0$, thereby proving that E/E_{tor} is torsion free. It is also finitely generated.

Assume now that M is a torsion free module which is finitely generated. Let $\{v_1, \dots, v_n\}$ be a maximal set of elements of M among a given finite set of generators $\{y_1, \dots, y_m\}$ such that $\{v_1, \dots, v_n\}$ is linearly independent. If y is one of the generators, there exist elements $a, b_1, \dots, b_n \in R$ not all 0, such that

$$ay + b_1 v_1 + \cdots + b_n v_n = 0.$$

Then $a \neq 0$ (otherwise we contradict the linear independence of $v_1, \dots, v_n$). Hence ay lies in $(v_1, \dots, v_n)$. Thus for each $j = 1, \dots, m$ we can find $a_j \in R$, $a_j \neq 0$, such that $a_j y_j$ lies in $(v_1, \dots, v_n)$. Let $a = a_1 \cdots a_m$ be the product. Then aM is contained in $(v_1, \dots, v_n)$, and $a \neq 0$. The map

$$x \mapsto ax$$

is an injective homomorphism, whose image is contained in a free module. This image is isomorphic to M, and we conclude from Theorem 7.1 that M is free, as desired.

To get the submodule F we need a lemma.

Lemma 7.4. *Let E, E' be modules, and assume that E' is free. Let $f: E \to E'$ be a surjective homomorphism. Then there exists a free submodule F of E such that the restriction of f to F induces an isomorphism of F with E', and such that $E = F \oplus \operatorname{Ker} f$.*

Proof. Let $\{x_i'\}_{i \in I}$ be a basis of E'. For each i, let x_i be an element of E such that $f(x_i) = x_i'$. Let F be the submodule of E generated by all the elements x_i, $i \in I$. Then one sees at once that the family of elements $\{x_i\}_{i \in I}$ is linearly independent, and therefore that F is free. Given $x \in E$, there exist elements $a_i \in R$ such that

$$f(x) = \sum a_i x_i'.$$

Then $x - \sum a_i x_i$ lies in the kernel of f, and therefore $E = \operatorname{Ker} f + F$. It is clear that $\operatorname{Ker} f \cap F = 0$, and hence that the sum is direct, thereby proving the lemma.

We apply the lemma to the homomorphism $E \to E/E_{\text{tor}}$ in Theorem 7.3 to get our decomposition $E = E_{\text{tor}} \oplus F$. The dimension of F is uniquely determined, because F is isomorphic to E/E_{tor} for any decomposition of E into a direct sum as stated in the theorem.

The dimension of the free module F in Theorem 7.3 is called the **rank** of E.

In order to get the structure theorem for finitely generated modules over R, one can proceed exactly as for abelian groups. We shall describe the dictionary which allows us to transport the proofs essentially without change.

Let E be a module over R. Let $x \in E$. The map $a \mapsto ax$ is a homomorphism of R onto the submodule generated by x, and the kernel is an ideal, which is principal, generated by an element $m \in R$. We say that m is a **period** of x. We

note that m is determined up to multiplication by a unit (if $m \neq 0$). An element $c \in R$, $c \neq 0$, is said to be an **exponent** for E (resp. for x) if $cE = 0$ (resp. $cx = 0$).

Let p be a prime element. We denote by $E(p)$ the submodule of E consisting of all elements x having an exponent which is a power p^r ($r \geqq 1$). A p-submodule of E is a submodule contained in $E(p)$.

We select once and for all a system of representatives for the prime elements of R (modulo units). For instance, if R is a polynomial ring in one variable over a field, we take as representatives the irreducible polynomials with leading coefficient 1.

Let $m \in R$, $m \neq 0$. We denote by E_m the kernel of the map $x \mapsto mx$. It consists of all elements of E having exponent m.

A module E is said to be **cyclic** if it is isomorphic to $R/(a)$ for some element $a \in R$. Without loss of generality if $a \neq 0$, one may assume that a is a product of primes in our system of representatives, and then we could say that a is the order of the module.

Let $r_1, \ldots, r_s$ be integers $\geqq 1$. A p-module E is said to be of **type**

$$(p^{r_1}, \ldots, p^{r_s})$$

if it is isomorphic to the product of cyclic modules $R/(p^{r_i})$ ($i = 1, \ldots, s$). If p is fixed, then one could say that the module is of type $(r_1, \ldots, r_s)$ (relative to p).

All the proofs of Chapter I, §8 now go over without change. Whenever we argue on the size of a positive integer m, we have a similar argument on the number of prime factors appearing in its prime factorization. If we deal with a prime power p^r, we can view the order as being determined by r. The reader can now check that the proofs of Chapter I, §8 are applicable.

However, we shall develop the theory once again without assuming any knowledge of Chapter I, §8. Thus our treatment is self-contained.

Theorem 7.5. *Let E be a finitely generated torsion module $\neq 0$. Then E is the direct sum*

$$E = \bigoplus_p E(p),$$

taken over all primes p such that $E(p) \neq 0$. Each $E(p)$ can be written as a direct sum

$$E(p) = R/(p^{\nu_1}) \oplus \cdots \oplus R/(p^{\nu_s})$$

with $1 \leqq \nu_1 \leqq \cdots \leqq \nu_s$. The sequence $\nu_1, \ldots, \nu_s$ is uniquely determined.

Proof. Let a be an exponent for E, and suppose that $a = bc$ with $(b, c) = (1)$. Let $x, y \in R$ be such that

$$1 = xb + yc.$$

We contend that $E = E_b \oplus E_c$. Our first assertion then follows by induction, expressing a as a product of prime powers. Let $v \in E$. Then

$$v = xbv + ycv.$$

Then $xbv \in E_c$ because $cxbv = xav = 0$. Similarly, $ycv \in E_b$. Finally $E_b \cap E_c = 0$, as one sees immediately. Hence E is the direct sum of E_b and E_c.

We must now prove that $E(p)$ is a direct sum as stated. If $y_1, \dots, y_m$ are elements of a module, we shall say that they are **independent** if whenever we have a relation

$$a_1 y_1 + \cdots + a_m y_m = 0$$

with $a_i \in R$, then we must have $a_i y_i = 0$ for all i. (Observe that **independent** does not mean **linearly independent**.) We see at once that $y_1, \dots, y_m$ are independent if and only if the module $(y_1, \dots, y_m)$ has the direct sum decomposition

$$(y_1, \dots, y_m) = (y_1) \oplus \cdots \oplus (y_m)$$

in terms of the cyclic modules (y_i), $i = 1, \dots, m$.

We now have an analogue of Lemma 7.4 for modules having a prime power exponent.

Lemma 7.6. *Let E be a torsion module of exponent p^r $(r \geqq 1)$ for some prime element p. Let $x_1 \in E$ be an element of period p^r. Let $\bar{E} = E/(x_1)$. Let $\bar{y}_1, \dots, \bar{y}_m$ be independent elements of $\bar{E}$. Then for each i there exists a representative $y_i \in E$ of $\bar{y}_i$, such that the period of y_i is the same as the period of $\bar{y}_i$. The elements $x_1, y_1, \dots, y_m$ are independent.*

Proof. Let $\bar{y} \in \bar{E}$ have period p^n for some $n \geqq 1$. Let y be a representative of $\bar{y}$ in E. Then $p^n y \in (x_1)$, and hence

$$p^n y = p^s c x_1, \qquad c \in R, p \nmid c,$$

for some $s \leqq r$. If $s = r$, we see that y has the same period as $\bar{y}$. If $s < r$, then $p^s c x_1$ has period p^{r-s}, and hence y has period p^{n+r-s}. We must have

$$n + r - s \leqq r,$$

because p^r is an exponent for E. Thus we obtain $n \leqq s$, and we see that

$$y - p^{s-n} c x_1$$

is a representative for $\bar{y}$, whose period is p^n.

Let y_i be a representative for $\bar{y}_i$ having the same period. We prove that $x_1, y_1, \dots, y_m$ are independent. Suppose that $a, a_1, \dots, a_m \in R$ are elements such that

$$a x_1 + a_1 y_1 + \cdots + a_m y_m = 0.$$

Then
$$a_1\bar{y}_1 + \cdots + a_m\bar{y}_m = 0.$$

By hypothesis, we must have $a_i\bar{y}_i = 0$ for each i. If p^{r_i} is the period of $\bar{y}_i$, then p^{r_i} divides a_i. We then conclude that $a_iy_i = 0$ for each i, and hence finally that $ax_1 = 0$, thereby proving the desired independence.

To get the direct sum decomposition of $E(p)$, we first note that $E(p)$ is finitely generated. We may assume without loss of generality that $E = E(p)$. Let x_1 be an element of E whose period p^{r_1} is such that r_1 is maximal. Let $\bar{E} = E/(x_1)$. We contend that dim $\bar{E}_p$ as vector space over R/pR is strictly less than dim E_p. Indeed, if $\bar{y}_1, \ldots, \bar{y}_m$ are linearly independent elements of $\bar{E}_p$ over R/pR, then Lemma 7.6 implies that $\dim E_p \geqq m + 1$ because we can always find an element of (x_1) having period p, independent of $y_1, \ldots, y_m$. Hence $\dim \bar{E}_p < \dim E_p$. We can prove the direct sum decomposition by induction. If $E \neq 0$, there exist elements $\bar{x}_2, \ldots, \bar{x}_s$ having periods $p^{r_2}, \ldots, p^{r_s}$ respectively, such that $r_2 \geqq \cdots \geqq r_s$. By Lemma 7.6, there exist representatives $x_2, \ldots, x_r$ in E such that x_i has period p^{r_i} and $x_1, \ldots, x_r$ are independent. Since p^{r_1} is such that r_1 is maximal, we have $r_1 \geqq r_2$, and our decomposition is achieved.

The uniqueness will be a consequence of a more general uniqueness theorem, which we state next.

Theorem 7.7. *Let E be a finitely generated torsion module, $E \neq 0$. Then E is isomorphic to a direct sum of non-zero factors*

$$R/(q_1) \oplus \cdots \oplus R/(q_r),$$

where $q_1, \ldots, q_r$ are non-zero elements of R, and $q_1 | q_2 | \cdots | q_r$. The sequence of ideals $(q_1), \ldots, (q_r)$ is uniquely determined by the above conditions.

Proof. Using Theorem 7.5, decompose E into a direct sum of p-submodules, say $E(p_1) \oplus \cdots \oplus E(p_l)$, and then decompose each $E(p_i)$ into a direct sum of cyclic submodules of periods $p_i^{r_{ij}}$. We visualize these symbolically as described by the following diagram:

$$\begin{array}{llll} E(p_1): & r_{11} \leqq & r_{12} \leqq & \cdots \\ E(p_2): & r_{21} \leqq & r_{22} \leqq & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ E(p_l): & r_{l1} \leqq & r_{l2} \leqq & \cdots \end{array}$$

A horizontal row describes the type of the module with respect to the prime at the left. The exponents r_{ij} are arranged in increasing order for each fixed $i = 1, \ldots, l$. We let $q_1, \ldots, q_r$ correspond to the columns of the matrix of exponents, in other words

$$q_1 = p_1^{r_{11}}p_2^{r_{21}} \cdots p_l^{r_{l1}},$$

$$q_2 = p_1^{r_{12}}p_2^{r_{22}} \cdots p_l^{r_{l2}},$$

$$\cdots$$

The direct sum of the cyclic modules represented by the first column is then isomorphic to $R/(q_1)$, because, as with abelian groups, the direct sum of cyclic modules whose periods are relatively prime is also cyclic. We have a similar remark for each column, and we observe that our proof actually orders the q_j by increasing divisibility, as was to be shown.

Now for uniqueness. Let p be any prime, and suppose that $E = R/(pb)$ for some $b \in R$, $b \neq 0$. Then E_p is the submodule $bR/(pb)$, as follows at once from unique factorization in R. But the kernel of the composite map

$$R \to bR \to bR/(pb)$$

is precisely (p). Thus we have an isomorphism

$$R/(p) \approx bR/(pb).$$

Let now E be expressed as in the theorem, as a direct sum of r terms. An element

$$v = v_1 \oplus \cdots \oplus v_r, \qquad v_i \in R/(q_i)$$

is in E_p if and only if $pv_i = 0$ for all i. Hence E_p is the direct sum of the kernel of multiplication by p in each term. But E_p is a vector space over $R/(p)$, and its dimension is therefore equal to the number of terms $R/(q_i)$ such that p divides q_i.

Suppose that p is a prime dividing q_1, and hence q_i for each $i = 1, \ldots, r$. Let E have a direct sum decomposition into d terms satisfying the conditions of the theorem, say

$$E = R/(q'_1) \oplus \cdots \oplus R/(q'_s).$$

Then p must divide at least r of the elements q'_j, whence $r \leqq s$. By symmetry, $r = s$, and p divides q'_j for all j.

Consider the module pE. By a preceding remark, if we write $q_i = pb_i$, then

$$pE \approx R/(b_1) \oplus \cdots \oplus R/(b_r),$$

and $b_1 | \cdots | b_r$. Some of the b_i may be units, but those which are not units determine their principal ideal uniquely, by induction. Hence if

$$(b_1) = \cdots = (b_j) = 1$$

but $(b_{j+1}) \neq (1)$, then the sequence of ideals

$$(b_{j+1}), \ldots, (b_r)$$

is uniquely determined. This proves our uniqueness statement, and concludes the proof of Theorem 7.7.

The ideals $(q_1), \dots, (q_r)$ are called the **invariants** of E.

For one of the main applications of Theorem 7.7 to linear algebra, see Chapter XV, §2.

The next theorem is included for completeness. It is called the **elementary divisors** theorem.

Theorem 7.8. *Let F be a free module over R, and let M be a finitely generated submodule $\neq 0$. Then there exists a basis $\mathcal{B}$ of F, elements $e_1, \dots, e_m$ in this basis, and non-zero elements $a_1, \dots, a_m \in R$ such that:*

(i) *The elements $a_1e_1, \dots, a_me_m$ form a basis of M over R.*

(ii) *We have $a_i | a_{i+1}$ for $i = 1, \dots, m - 1$.*

The sequence of ideals $(a_1), \dots, (a_m)$ is uniquely determined by the preceding conditions.

Proof. Write a finite set of generators for M as linear combination of a finite number of elements in a basis for F. These elements generate a free submodule of finite rank, and thus it suffices to prove the theorem when F has finite rank, which we now assume. We let $n = \text{rank}(F)$.

The uniqueness is a corollary of Theorem 7.7. Suppose we have a basis as in the theorem. Say $a_1, \dots, a_s$ are units, and so can be taken to be $= 1$, and $a_{s+j} = q_j$ with $q_1 | q_2 | \dots | q_r$ non-units. Observe that $F/M = \bar{F}$ is a finitely generated module over R, having the direct sum expression

$$F/M = \bar{F} \approx \bigoplus_{j=1}^{r} (R/q_jR)\bar{e}_j \oplus \text{free module of rank } n - (r + s)$$

where a bar denotes the class of an element of F mod M. Thus the direct sum over $j = 1, \dots, r$ is the torsion submodule of $\bar{F}$, whence the elements $q_1, \dots, q_r$ are uniquely determined by Theorem 7.7. We have $r + s = m$, so the rank of F/M is $n - m$, which determines m uniquely. Then $s = m - r$ is uniquely determined as the number of units among $a_1, \dots, a_m$. This proves the uniqueness part of the theorem. Next we prove existence.

Let λ be a functional on F, in other words, an element of $\text{Hom}_R(F, R)$. We let $J_\lambda = \lambda(M)$. Then J_λ is an ideal of R. Select λ_1 such that $\lambda_1(M)$ is maximal in the set of ideals $\{J_\lambda\}$, that is to say, there is no properly larger ideal in the set $\{J_\lambda\}$.

Let $\lambda_1(M) = (a_1)$. Then $a_1 \neq 0$, because there exists a non-zero element of M, and expressing this element in terms of some basis for F over R, with some non-zero coordinate, we take the projection on this coordinate to get a functional whose value on M is not 0. Let $x_1 \in M$ be such that $\lambda_1(x_1) = a_1$. For any functional g we must have $g(x_1) \in (a_1)$ [immediate from the maximality of

$\lambda_1(M)]$. Writing x_1 in terms of any basis of F, we see that its coefficients must all be divisible by a_1. (If some coefficient is not divisible by a_1, project on this coefficient to get an impossible functional.) Therefore we can write $x_1 = a_1 e_1$ with some element $e_1 \in F$.

Next we prove that F is a direct sum

$$F = Re_1 \oplus \operatorname{Ker} \lambda_1.$$

Since $\lambda_1(e_1) = 1$, it is clear that $Re_1 \cap \operatorname{Ker} \lambda_1 = 0$. Furthermore, given $x \in F$ we note that $x - \lambda_1(x)e_1$ is in the kernel of λ_1. Hence F is the sum of the indicated submodules, and therefore the direct sum.

We note that $\operatorname{Ker} \lambda_1$ is free, being a submodule of a free module (Theorem 7.1). We let

$$F_1 = \operatorname{Ker} \lambda_1 \quad \text{and} \quad M_1 = M \cap \operatorname{Ker} \lambda_1.$$

We see at once that $\quad M = Rx_1 \oplus M_1.$

Thus M_1 is a submodule of F_1 and its dimension is one less than the dimension of M. From the maximality condition on $\lambda_1(M)$, it follows at once that for any functional λ on F_1, the image $\lambda(M)$ will be contained in $\lambda_1(M)$ (because otherwise, a suitable linear combination of functionals would yield an ideal larger than (a_1)). We can therefore complete the existence proof by induction.

In Theorem 7.8, we call the ideals $(a_1), \ldots, (a_m)$ the **invariants** of M in F. For another characterization of these invariants, see Chapter XIII, Proposition 4.20.

Example. First, see examples of situations similar to those of Theorem 7.8 in Exercises 5, 7, and 8, and for Dedekind rings in Exercise 13.

Example. Another way to obtain a module M as in Theorem 7.8 is as a module of relations. Let W be a finitely generated module over R, with generators $w_1, \ldots, w_n$. By a **relation** among $\{w_1, \ldots, w_n\}$ we mean an element $(a_1, \ldots, a_n) \in R^n$ such that $\sum a_i w_i = 0$. The set of such relations is a submodule of R^n, to which Theorem 7.8 may be applied.

It is also possible to formulate a proof of Theorem 7.8 by considering M as a submodule of R^n, and applying the method of row and column operations to get a desired basis. In this context, we make some further comments which may serve to illustrate Theorem 7.8. We assume that the reader is acquainted with matrices over a ring. By **row operations** we mean: interchanging two rows; adding a multiple of one row to another; multiplying a row by a unit in the ring. We define **column operations** similarly. These row and column operations correspond to multiplication with the so-called elementary matrices in the ring.

Theorem 7.9. *Assume that the elementary matrices in R generate $GL_n(R)$. Let (x_{ij}) be a non-zero matrix with components in R. Then with a finite number of row and column operations, it is possible to bring the matrix to the form*

$$\begin{pmatrix} a_1 & 0 & \cdots & \cdot & \cdot & \cdots & 0 \\ 0 & a_2 & \cdots & \cdot & \cdot & \cdots & 0 \\ \vdots & & \ddots & & & & \vdots \\ 0 & \cdot & \cdots & a_m & \cdot & \cdots & 0 \\ 0 & \cdot & \cdots & \cdot & 0 & \cdots & 0 \\ \vdots & & & & & & \vdots \\ 0 & \cdot & \cdots & \cdot & \cdot & \cdots & 0 \end{pmatrix}$$

with $a_1 \cdots a_m \neq 0$ *and* $a_1 | a_2 | \cdots | a_m$.

We leave the proof for the reader. Either Theorem 7.9 can be viewed as equivalent to Theorem 7.8, or a direct proof may be given. In any case, Theorem 7.9 can be used in the following context. Consider a system of linear equations

$$c_{11}x_1 + \cdots + c_{1n}x_n = 0$$

$$\cdots$$

$$c_{r1}x_1 + \cdots + c_{rn}x_n = 0.$$

with coefficients in R. Let F be the submodule of R^n generated by the vectors $X = (x_1, \ldots, x_n)$ which are solutions of this system. By Theorem 7.1, we know that F is free of dimension $\leqq n$. Theorem 7.9 can be viewed as providing a normalized basis for F in line with Theorem 7.8.

Further example. As pointed out by Paul Cohen, the row and column method can be applied to modules over a power series ring $\mathfrak{o}[[X]]$, where $\mathfrak{o}$ is a complete discrete valuation ring. Cf. Theorem 3.1 of Chapter 5 in my *Cyclotomic Fields* I *and* II (Springer Verlag, 1990). For instance, one could pick $\mathfrak{o}$ itself to be a power series ring $k[[T]]$ in one variable over a field k, but in the theory of cyclotomic fields in the above reference, $\mathfrak{o}$ is taken to be the ring of p-adic integers. On the other hand, George Bergman has drawn my attention to P. M. Cohn's "On the structure of GL_2 of a ring," *IHES Publ. Math.* No. 30 (1966), giving examples of principal rings where one cannot use row and column operations in Theorem 7.9.

§8. EULER-POINCARÉ MAPS

The present section may be viewed as providing an example and application of the Jordan-Hölder theorem for modules. But as pointed out in the examples and references below, it also provides an introduction for further theories.

Again let A be a ring. We continue to consider A-modules. Let Γ be an abelian group, written additively. Let φ be a rule which to certain modules associates an element of Γ, subject to the following condition:

If $0 \to M' \to M \to M'' \to 0$ *is exact, then* $\varphi(M)$ *is defined if and only if* $\varphi(M')$ *and* $\varphi(M'')$ *are defined, and in that case, we have*

$$\varphi(M) = \varphi(M') + \varphi(M'').$$

Furthermore $\varphi(0)$ *is defined and equal to* 0.

Such a rule φ will be called an **Euler-Poincaré mapping** on the category of A-modules. If M' is isomorphic to M, then from the exact sequence

$$0 \to M' \to M \to 0 \to 0$$

we conclude that $\varphi(M')$ is defined if $\varphi(M)$ is defined, and that $\varphi(M') = \varphi(M)$. Thus if $\varphi(M)$ is defined for a module M, φ is defined on every submodule and factor module of M. In particular, if we have an exact sequence of modules

$$M' \to M \to M''$$

and if $\varphi(M')$ and $\varphi(M'')$ are defined, then so is $\varphi(M)$, as one sees at once by considering the kernel and image of our two maps, and using the definition.

Examples. We could let $A = \mathbf{Z}$, and let φ be defined for all finite abelian groups, and be equal to the order of the group. The value of φ is in the multiplicative group of positive rational numbers.

As another example, we consider the category of vector spaces over a field k. We let φ be defined for finite dimensional spaces, and be equal to the dimension. The values of φ are then in the additive group of integers.

In Chapter XV we shall see that the characteristic polynomial may be considered as an Euler-Poincaré map.

Observe that the natural map of a finite module into its image in the Grothendieck group defined at the end of §4 is a universal Euler-Poincaré mapping. We shall develop a more extensive theory of this mapping in Chapter XX, §3.

If M is a module (over a ring A), then a sequence of submodules

$$M = M_1 \supset M_2 \supset \cdots \supset M_r = 0$$

is also called a **finite filtration**, and we call r the **length** of the filtration. A module M is said to be **simple** if it does not contain any submodule other than 0 and M itself, and if $M \neq 0$. A filtration is said to be **simple** if each M_i/M_{i+1} is simple. *The Jordan-Hölder theorem asserts that two simple filtrations of a module are equivalent.*

A module M is said to be of **finite length** if it is 0 or if it admits a simple (finite) filtration. By the Jordan-Hölder theorem, the length of such a simple filtration is the uniquely determined, and is called the **length of the module**. In the language of Euler characteristics, the Jordan-Hölder theorem can be reformulated as follows:

Theorem 8.1. *Let φ be a rule which to each simple module associates an element of a commutative group Γ, and such that if $M \approx M'$ then*

$$\varphi(M) = \varphi(M').$$

Then φ has a unique extension to an Euler-Poincaré mapping defined on all modules of finite length.

Proof. Given a simple filtration

$$M = M_1 \supset M_2 \supset \cdots \supset M_r = 0$$

we define

$$\varphi(M) = \sum_{i=1}^{r-1} \varphi(M_i/M_{i+1}).$$

The Jordan-Hölder theorem shows immediately that this is well-defined, and that this extension of φ is an Euler-Poincaré map.

In particular, we see that the length function is the Euler-Poincaré map taking its values in the additive group of integers, and having the value 1 for any simple module.

§9. THE SNAKE LEMMA

This section gives a very general lemma, which will be used many times, so we extract it here. The reader may skip it until it is encountered, but already we give some exercises which show how it is applied: the five lemma in Exercise 15 and also Exercise 26. Other substantial applications in this book will occur in Chapter XVI, §3 in connection with the tensor product, and in Chapter XX in connection with complexes, resolutions, and derived functors.

We begin with routine comments. Consider a commutative diagram of homomorphisms of modules.

$$\begin{array}{ccc} M' & \xrightarrow{f} & M \\ {\scriptstyle d'}\downarrow & & \downarrow{\scriptstyle d} \\ N' & \xrightarrow[h]{} & N \end{array}$$

Then f induces a homomorphism

$$\operatorname{Ker} d' \to \operatorname{Ker} d.$$

Indeed, suppose $d'x' = 0$. Then $df(x') = 0$ because $df(x') = hd'(x') = 0$.

Similarly, h induces a homomorphism

$$\text{Coker } d' \to \text{Coker } d$$

in a natural way as follows. Let $y' \in N'$ represent an element of $N'/d'M'$. Then hy' mod dM does not depend on the choice of y' representing the given element, because if $y'' = y' + d'x'$, then

$$hy'' = hy' + hd'x' = hy' + dfx' \equiv hy' \bmod dM.$$

Thus we get a map

$$h_*: N'/d'M' = \text{Coker } d' \to N/dM = \text{Coker } d,$$

which is immediately verified to be a homomorphism.

In practice, given a commutative diagram as above, one sometimes writes f instead of h, so one writes f for the horizontal maps both above and below the diagram. This simplifies the notation, and is not so incorrect: we may view M', N' as the two components of a direct sum, and similarly for M, N. Then f is merely a homomorphism defined on the direct sum $M' \oplus N'$ into $M \oplus N$.

The snake lemma concerns a commutative and exact diagram called a **snake diagram**:

$$\begin{array}{ccccccc} & & M' & \xrightarrow{f} & M & \xrightarrow{g} & M'' & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle d'} & & \downarrow{\scriptstyle d} & & \downarrow{\scriptstyle d''} & & \\ 0 & \longrightarrow & N' & \xrightarrow[f]{} & N & \xrightarrow[g]{} & N'' & & \end{array}$$

Let $z'' \in \text{Ker } d''$. We can construct elements of N' as follows. Since g is surjective, there exists an element $z \in M$ such that $gz = z''$. We now move vertically down by d, and take dz. The commutativity $d''g = gd$ shows that $gdz = 0$ whence dz is in the kernel of g in N. By exactness, there exists an element $z' \in N'$ such that $fz' = dz$. In brief, we write

$$z' = f^{-1} \circ d \circ g^{-1} z''.$$

Of course, z' is not well defined because of the choices made when taking inverse images. However, the snake lemma will state exactly what goes on.

Lemma 9.1. (Snake Lemma). *Given a snake diagram as above, the map*

$$\delta : \text{Ker } d'' \to \text{Coker } d'$$

given by $\delta z'' = f^{-1} \circ d \circ g^{-1} z''$ *is well defined, and we have an exact sequence*

$$\text{Ker } d' \to \text{Ker } d \to \text{Ker } d'' \xrightarrow{\delta} \text{Coker } d' \to \text{Coker } d \to \text{Coker } d''$$

where the maps besides δ *are the natural ones.*

Proof. It is a routine verification that the class of z' mod Im d' is independent of the choices made when taking inverse images, whence defining the map δ. The proof of the exactness of the sequence is then routine, and consists in chasing around diagrams. It should be carried out in full detail by the reader who wishes to acquire a feeling for this type of triviality. As an example, we shall prove that

$$\operatorname{Ker} \delta \subset \operatorname{Im} g_*$$

where g_* is the induced map on kernels. Suppose the image of z'' is 0 in Coker d'. By definition, there exists $u' \in M'$ such that $z' = d'u'$. Then

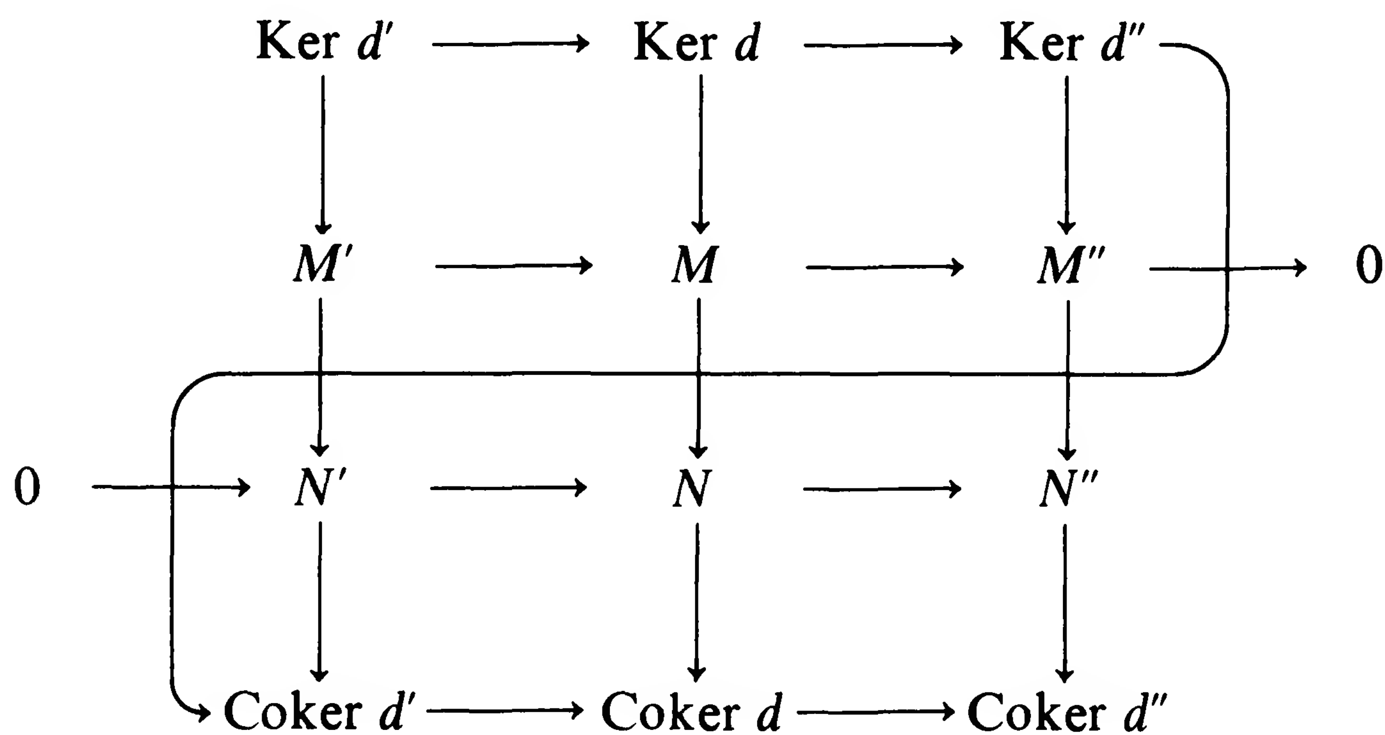
$$dz = fz' = fd'u' = dfu'$$

by commutativity. Hence

$$d(z - fu') = 0,$$

and $z - fu'$ is in the kernel of d. But $g(z - fu') = gz = z''$. This means that z'' is in the image of g_*, as desired. All the remaining cases of exactness will be left to the reader.

The original snake diagram may be completed by writing in the kernels and cokernels as follows (whence the name of the lemma):



§10. DIRECT AND INVERSE LIMITS

We return to limits, which we considered for groups in Chapter I. We now consider limits in other categories (rings, modules), and we point out that limits satisfy a universal property, in line with Chapter I, §11.

Let $I = \{i\}$ be a directed system of indices, defined in Chapter I, §10. Let $\mathfrak{A}$ be a category, and $\{A_i\}$ a family of objects in $\mathfrak{A}$. For each pair i, j such that

$i \leqq j$ assume given a morphism

$$f_j^i : A_i \to A_j$$

such that, whenever $i \leqq j \leqq k$, we have

$$f_k^j \circ f_j^i = f_k^i \quad \text{and} \quad f_i^i = \text{id}.$$

Such a family will be called a **directed family of morphisms.** A **direct limit** for the family $\{f_j^i\}$ is a universal object in the following category $\mathcal{C}$. $\text{Ob}(\mathcal{C})$ consists of pairs $(A, (f^i))$ where $A \in \text{Ob}(\mathfrak{A})$ and (f^i) is a family of morphisms $f^i : A_i \to A$, $i \in I$, such that for all $i \leqq j$ the following diagram is commutative:

$$\begin{array}{ccc} A_i & \xrightarrow{f_j^i} & A_j \\ & {\scriptstyle f^i}\searrow \quad \swarrow{\scriptstyle f^j} & \\ & A & \end{array}$$

(Universal of course means universally repelling.)

Thus if $(A, (f^i))$ is the direct limit, and if $(B, (g^i))$ is any object in the above category, then there exists a unique morphism $\varphi : A \to B$ which makes the following diagram commutative:

$$\begin{array}{ccc} A_i & \xrightarrow{f_j^i} & A_j \\ & {\scriptstyle f^i}\searrow \quad \swarrow{\scriptstyle f^j} & \\ & A & \\ {\scriptstyle g^i} & \downarrow{\scriptstyle \varphi} & {\scriptstyle g^j} \\ & B & \end{array}$$

For simplicity, one usually writes

$$A = \varinjlim_i A_i,$$

omitting the f_j^i from the notation.

Theorem 10.1. *Direct limits exist in the category of abelian groups, or more generally in the category of modules over a ring.*

Proof. Let $\{M_i\}$ be a directed system of modules over a ring. Let M be their direct sum. Let N be the submodule generated by all elements

$$x_{ij} = (\ldots, 0, x, 0, \ldots, -f_j^i(x), 0, \ldots)$$

where, for a given pair of indices (i, j) with $j \geqq i$, x_{ij} has component x in M_i, $f^i_j(x)$ in M_j, and component 0 elsewhere. Then we leave to the reader the verification that the factor module M/N is a direct limit, where the maps of M_i into M/N are the natural ones arising from the composite homomorphism

$$M_i \to M \to M/N.$$

Example. Let X be a topological space, and let $x \in X$. The open neighborhoods of x form a directed system, by inclusion. Indeed, given two open neighborhoods U and V, then $U \cap V$ is also an open neighborhood contained in both U and V. In sheaf theory, one assigns to each U an abelian group $A(U)$ and for each pair $U \supset V$ a homomorphism $h^U_V : A(U) \to A(V)$ such that if $U \supset V \supset W$ then $h^V_W \circ h^U_V = h^U_W$. Then the family of such homomorphisms is a directed family. The direct limit

$$\varinjlim_{U} A(U)$$

is called the **stalk** at the point x. We shall give the formal definition of a sheaf of abelian groups in Chapter XX, §6. For further reading, I recommend at least two references. First, the self-contained short version of Chapter II in Hartshorne's *Algebraic Geometry*, Springer Verlag, 1977. (Do all the exercises of that section, concerning sheaves.) The section is only five pages long. Second, I recommend the treatment in Gunning's *Introduction to Holomorphic Functions of Several Variables*, Wadsworth and Brooks/Cole, 1990.

We now reverse the arrows to define inverse limits. We are again given a directed set I and a family of objects A_i. If $j \geqq i$ we are now given a morphism

$$f^j_i : A_j \to A_i$$

satisfying the relations

$$f^i_k \circ f^j_i = f^j_k \quad \text{and} \quad f^i_i = \text{id},$$

if $j \geqq i$ and $i \geqq k$. As in the direct case, we can define a category of objects (A, f_i) with $f_i : A \to A_i$ such that for all i, j the following diagram is commutative:

$$\begin{array}{ccc} & A & \\ {\scriptstyle f_j}\swarrow & & \searrow{\scriptstyle f_i} \\ A_j & \xrightarrow[f^j_i]{} & A_i \end{array}$$

A universal object in this category is called an **inverse limit** of the system (A_i, f^i_j).

As before, we often say that

$$A = \varprojlim_i A_i$$

is the inverse limit, omitting the f^i_j from the notation.

Theorem 10.2. *Inverse limits exist in the category of groups, in the category of modules over a ring, and also in the category of rings.*

Proof. Let $\{G_i\}$ be a directed family of groups, for instance, and let Γ be their inverse limit as defined in Chapter I, §10. Let $p_i: \Gamma \to G_i$ be the projection (defined as the restriction from the projection of the direct product, since Γ is a subgroup of $\prod G_i$). It is routine to verify that these data give an inverse limit in the category of groups. The same construction also applies to the category of rings and modules.

Example. Let p be a prime number. For $n \geqq m$ we have a canonical surjective ring homomorphism

$$f^n_m: \mathbf{Z}/p^n\mathbf{Z} \to \mathbf{Z}/p^m\mathbf{Z}.$$

The projective limit is called the ring of ***p*-adic integers**, and is denoted by $\mathbf{Z}_p$. For a consideration of this ring as a complete discrete valuation ring, see Exercise 17 and Chapter XII.

Let k be a field. The power series ring $k[[T]]$ in one variable may be viewed as the projective limit of the factor polynomial rings $k[T]/(T^n)$, where for $n \geqq m$ we have the canonical ring homomorphism

$$f^n_m: k[T]/(T^n) \to k[T]/(T^m).$$

A similar remark applies to power series in several variables.

More generally, let R be a commutative ring and let J be a proper ideal. If $n \geqq m$ we have the canonical ring homomorphism

$$f^n_m: R/J^n \to R/J^m.$$

Let $\bar{R}_J = \lim R/J^n$ be the projective limit. Then R has a natural homomorphism into $\bar{R}_J$. If R is a Noetherian local ring, then by Krull's theorem (Theorem 5.6 of Chapter X), one knows that $\bigcap J^n = \{0\}$, and so the natural homorphism of R in its completion is an embedding. This construction is applied especially when J is the maximal ideal. It gives an algebraic version of the notion of holomorphic functions for the following reason.

Let R be a commutative ring and J a proper ideal. Define a ***J*-Cauchy sequence** $\{x_n\}$ to be a sequence of elements of R satisfying the following condition. Given a positive integer k there exists N such that for all n, $m \geqq N$ we have $x_n - x_m \in J^k$. Define a **null sequence** to be a sequence for which given k there exists N such that for all $n \geqq N$ we have $x_n \in J^k$. Define addition and multipli-

cation of sequences termwise. Then the Cauchy sequences form a ring $\mathcal{C}$, the null sequences form an ideal $\mathcal{N}$, and the factor ring $\mathcal{C}/\mathcal{N}$ is called the J-**adic completion** of R. Prove these statements as an exercise, and also prove that there is a natural isomorphism

$$\mathcal{C}/\mathcal{N} \approx \varprojlim R/J^n.$$

Thus the inverse limit $\varprojlim R/J^n$ is also called the J-adic completion. See Chapter XII for the completion in the context of absolute values on fields.

Examples. In certain situations one wants to determine whether there exist solutions of a system of a polynomial equation $f(X_1, \ldots, X_n) = 0$ with coefficients in a power series ring $k[T]$, say in one variable. One method is to consider the ring mod (T^N), in which case this equation amounts to a finite number of equations in the coefficients. A solution of $f(X) = 0$ is then viewed as an inverse limit of truncated solutions. For an early example of this method see [La 52], and for an extension to several variables [Ar 68].

[La 52] S. LANG, On quasi algebraic closure, *Ann of Math.* **55** (1952), pp. 373-390

[Ar 68] M. ARTIN, On the solutions of analytic equations, *Invent. Math.* **5** (1968), pp. 277-291

See also Chapter XII, §7.

In Iwasawa theory, one considers a sequence of Galois cyclic extensions K_n over a number field k of degree p^n with p prime, and with $K_n \subset K_{n+1}$. Let G_n be the Galois group of K_n over k. Then one takes the inverse limit of the group rings $(\mathbf{Z}/p^n\mathbf{Z})[G_n]$, following Iwasawa and Serre. Cf. my *Cyclotomic Fields*, Chapter 5. In such towers of fields, one can also consider the projective limits of the modules mentioned as examples at the end of §1. Specifically, consider the group of p^n-th roots of unity $\boldsymbol{\mu}_{p^n}$, and let $K_n = \mathbf{Q}(\boldsymbol{\mu}_{p^{n+1}})$, with $K_0 = \mathbf{Q}(\boldsymbol{\mu}_p)$. We let

$$T_p(\boldsymbol{\mu}) = \varprojlim \boldsymbol{\mu}_{p^n}$$

under the homomorphisms $\boldsymbol{\mu}_{p^{n+1}} \to \boldsymbol{\mu}_{p^n}$ given by $\zeta \mapsto \zeta^p$. Then $T_p(\boldsymbol{\mu})$ becomes a module for the projective limits of the group rings. Similarly, one can consider inverse limits for each one of the modules given in the examples at the end of §1. (See Exercise 18.) The determination of the structure of these inverse limits leads to fundamental problems in number theory and algebraic geometry.

After such examples from real life after basic algebra, we return to some general considerations about inverse limits.

Let $(A_i, f_i^j) = (A_i)$ and $(B_i, g_i^j) = (B_i)$ be two inverse systems of abelian groups indexed by the same indexing set. A **homomorphism** $(A_i) \to (B_i)$ is the obvious thing, namely a family of homomorphisms

$$h_i : A_i \to B_i$$

for each i which commute with the maps of the inverse systems:

$$\begin{array}{ccc} A_j & \xrightarrow{h_j} & B_j \\ \downarrow{\scriptstyle f_i^j} & & \downarrow{\scriptstyle g_i^j} \\ A_i & \xrightarrow[h_i]{} & B_i \end{array}$$

A sequence

$$0 \to (A_i) \to (B_i) \to (C_i) \to 0$$

is said to be **exact** if the corresponding sequence of groups is exact for each i.

Let (A_n) be an inverse system of sets, indexed for simplicity by the positive integers, with connecting maps

$$u_{m,n}: A_m \to A_n \quad \text{for} \quad m \geqq n.$$

We say that this system satisfies the **Mittag-Leffler condition ML** if for each n, the decreasing sequence $u_{m,n}(A_m)$ $(m \geqq n)$ stabilizes, i.e. is constant for m sufficiently large. This condition is satisfied when $u_{m,n}$ is surjective for all m, n.

We note that trivially, the inverse limit functor is left exact, in the sense that given an exact sequence

$$0 \to (A_n) \to (B_n) \to (C_n) \to 0$$

then

$$0 \to \varprojlim A_n \to \varprojlim B_n \to \varprojlim C_n$$

is exact.

Proposition 10.3. *Assume that (A_n) satisfies* **ML**. *Given an exact sequence*

$$0 \to (A_n) \to (B_n) \xrightarrow{g} (C_n) \to 0$$

of inverse systems, then

$$0 \to \varprojlim A_n \to \varprojlim B_n \to \varprojlim C_n \to 0$$

is exact.

Proof. The only point is to prove the surjectivity on the right. Let (c_n) be an element of the inverse limit. Then each inverse image $g^{-1}(c_n)$ is a coset of A_n, so in bijection with A_n. These inverse images form an inverse system, and the **ML** condition on (A_n) implies **ML** on $(g^{-1}(c_n))$. Let S_n be the stable subset

$$S_n = \bigcap_{m \geqq n} u^B_{m,n}(g^{-1}(c_m)).$$

Then the connecting maps in the inverse system (S_n) are surjective, and so there is an element (b_n) in the inverse limit. It is immediate that g maps this element on the given (c_n), thereby concluding the proof of the Proposition.

Proposition 10.4. *Let (C_n) be an inverse system of abelian groups satisfying* **ML**, *and let $(u_{m,n})$ be the system of connecting maps. Then we have an exact sequence*

$$0 \to \varprojlim C_n \to \prod C_n \xrightarrow{1-u} \prod C_n \to 0.$$

Proof. For each positive integer N we have an exact sequence with a finite product

$$0 \to \lim_{1 \leqq n \leqq N} C_n \to \prod_{n=1}^{N} C_n \xrightarrow{1-u} \prod_{n=1}^{N} C_n \to 0.$$

The map u is the natural one, whose effect on a vector is

$$(0, \ldots, 0, c_m, 0, \ldots, 0) \mapsto (0, \ldots, 0, u_{m,m-1}c_m, 0, \ldots, 0).$$

One sees immediately that the sequence is exact. The infinite products are inverse limits taken over N. The hypothesis implies at once that **ML** is satisfied for the inverse limit on the left, and we can therefore apply Proposition 10.3 to conclude the proof.

EXERCISES

1. Let V be a vector space over a field K, and let U, W be subspaces. Show that
$$\dim U + \dim W = \dim(U + W) + \dim(U \cap W).$$

2. Generalize the dimension statement of Theorem 5.2 to free modules over a commutative ring. [*Hint*: Recall how an analogous statement was proved for free abelian groups, and use a maximal ideal instead of a prime number.]

3. Let R be an entire ring containing a field k as a subring. Suppose that R is a finite dimensional vector space over k under the ring multiplication. Show that R is a field.

4. **Direct sums.**
   (a) Prove in detail that the conditions given in Proposition 3.2 for a sequence to split are equivalent. Show that a sequence $0 \to M' \xrightarrow{f} M \xrightarrow{g} M'' \to 0$ splits if and only if there exists a submodule N of M such that M is equal to the direct sum $\text{Im} f \oplus N$, and that if this is the case, then N is isomorphic to M''. Complete all the details of the proof of Proposition 3.2.

(b) Let E and $E_i (i = 1, \dots, m)$ be modules over a ring. Let $\varphi_i \colon E_i \to E$ and $\psi_i \colon E \to E_i$ be homomorphisms having the following properties:

$$\psi_i \circ \varphi_i = \text{id}, \qquad \psi_i \circ \varphi_j = 0 \qquad \text{if } i \neq j,$$

$$\sum_{i=1}^{m} \varphi_i \circ \psi_i = \text{id}.$$

Show that the map $x \mapsto (\psi_1 x, \dots, \psi_m x)$ is an isomorphism of E onto the direct product of the E_i $(i = 1, \dots, m)$, and that the map

$$(x_1, \dots, x_m) \mapsto \varphi_1 x_1 + \cdots + \varphi_m x_m$$

is an isomorphism of this direct product onto E.

Conversely, if E is equal to a direct product (or direct sum) of submodules E_i $(i = 1, \dots, m)$, if we let φ_i be the inclusion of E_i in E, and ψ_i the projection of E on E_i, then these maps satisfy the above-mentioned properties.

5. Let A be an additive subgroup of Euclidean space $\mathbf{R}^n$, and assume that in every bounded region of space, there is only a finite number of elements of A. Show that A is a free abelian group on $\leqq n$ generators. [*Hint*: Induction on the maximal number of linearly independent elements of A over $\mathbf{R}$. Let $v_1, \dots, v_m$ be a maximal set of such elements, and let A_0 be the subgroup of A contained in the $\mathbf{R}$-space generated by $v_1, \dots, v_{m-1}$. By induction, one may assume that any element of A_0 is a linear integral combination of $v_1, \dots, v_{m-1}$. Let S be the subset of elements $v \in A$ of the form $v = a_1 v_1 + \cdots + a_m v_m$ with real coefficients a_i satisfying

$$0 \leqq a_i < 1 \qquad \text{if } i = 1, \dots, m-1$$
$$0 \leqq a_m \leqq 1.$$

If v'_m is an element of S with the smallest $a_m \neq 0$, show that $\{v_1, \dots, v_{m-1}, v'_m\}$ is a basis of A over $\mathbf{Z}$.]

Note. The above exercise is applied in algebraic number theory to show that the group of units in the ring of integers of a number field modulo torsion is isomorphic to a lattice in a Euclidean space. See Exercise 4 of Chapter VII.

6. (Artin-Tate). Let G be a finite group operating on a finite set S. For $w \in S$, denote $1 \cdot w$ by $[w]$, so that we have the direct sum

$$\mathbf{Z}\langle S\rangle = \sum_{w \in S} \mathbf{Z}[w].$$

Define an action of G on $\mathbf{Z}\langle S\rangle$ by defining $\sigma[w] = [\sigma w]$ (for $w \in S$), and extending σ to $\mathbf{Z}\langle S\rangle$ by linearity. Let M be a subgroup of $\mathbf{Z}\langle S\rangle$ of rank $\#[S]$. Show that M has a $\mathbf{Z}$-basis $\{y_w\}_{w \in S}$ such that $\sigma y_w = y_{\sigma w}$ for all $w \in S$. (Cf. my *Algebraic Number Theory*, Chapter IX, §4, Theorem 1.)

7. Let M be a finitely generated abelian group. By a **seminorm** on M we mean a real-valued function $v \mapsto |v|$ satisfying the following properties:

$$|v| \geqq 0 \text{ for all } v \in M;$$
$$|nv| = |n|\,|v| \text{ for } n \in \mathbf{Z};$$
$$|v + w| \leqq |v| + |w| \text{ for all } v, w \in M.$$

By the **kernel** of the seminorm we mean the subset of elements v such that $|v| = 0$.

(a) Let M_0 be the kernel. Show that M_0 is a subgroup. If $M_0 = \{0\}$, then the seminorm is called a **norm**.

(b) Assume that M has rank r. Let $v_1, \ldots, v_r \in M$ be linearly independent over $\mathbf{Z}$ mod M_0. Prove that there exists a basis $\{w_1, \ldots, w_r\}$ of M/M_0 such that

$$|w_i| \leqq \sum_{j=1}^{i} |v_j|.$$

[*Hint*: An explicit version of the proof of Theorem 7.8 gives the result. Without loss of generality, we can asume $M_0 = \{0\}$. Let $M_1 = \langle v_1, \ldots, v_r \rangle$. Let d be the exponent of M/M_1. Then dM has a finite index in M_1. Let $n_{j,j}$ be the smallest positive integer such that there exist integers $n_{j,1}, \ldots, n_{j,j-1}$ satisfying

$$n_{j,1}v_1 + \cdots + n_{j,j}v_j = dw_j \text{ for some } w_j \in M.$$

Without loss of generality we may assume $0 \leqq n_{j,k} \leqq d - 1$. Then the elements $w_1, \ldots, w_r$ form the desired basis.]

8. Consider the multiplicative group $\mathbf{Q}^*$ of non-zero rational numbers. For a non-zero rational number $x = a/b$ with $a, b \in \mathbf{Z}$ and $(a, b) = 1$, define the **height**

$$h(x) = \log \max(|a|, |b|).$$

(a) Show that h defines a seminorm on $\mathbf{Q}^*$, whose kernel consists of ± 1 (the torsion group).

(b) Let M_1 be a finitely generated subgroup of $\mathbf{Q}^*$, generated by rational numbers $x_1, \ldots, x_m$. Let M be the subgroup of $\mathbf{Q}^*$ consisting of those elements x such that $x^s \in M_1$ for some positive integer s. Show that M is finitely generated, and using Exercise 7, find a bound for the seminorm of a set of generators of M in terms of the seminorms of $x_1, \ldots, x_m$.

Note. The above two exercises are applied in questions of diophantine approximation. See my Diophantine approximation on toruses, *Am. J. Math.* **86** (1964), pp. 521-533, and the discussion and references I give in *Encyclopedia of Mathematical Sciences, Number Theory III*, Springer Verlag, 1991, pp. 240-243.

Localization

9. (a) Let A be a commutative ring and let M be an A-module. Let S be a multiplicative subset of A. Define $S^{-1}M$ in a manner analogous to the one we used to define $S^{-1}A$, and show that $S^{-1}M$ is an $S^{-1}A$-module.

(b) If $0 \to M' \to M \to M'' \to 0$ is an exact sequence, show that the sequence $0 \to S^{-1}M' \to S^{-1}M \to S^{-1}M'' \to 0$ is exact.

10. (a) If $\mathfrak{p}$ is a prime ideal, and $S = A - \mathfrak{p}$ is the complement of $\mathfrak{p}$ in the ring A, then $S^{-1}M$ is denoted by $M_{\mathfrak{p}}$. Show that the natural map

$$M \to \prod M_{\mathfrak{p}}$$

of a module M into the direct product of all localizations $M_{\mathfrak{p}}$ where $\mathfrak{p}$ ranges over all *maximal* ideals, is injective.

(b) Show that a sequence $0 \to M' \to M \to M'' \to 0$ is exact if and only if the sequence $0 \to M'_{\mathfrak{p}} \to M_{\mathfrak{p}} \to M''_{\mathfrak{p}} \to 0$ is exact for all primes $\mathfrak{p}$.

(c) Let A be an entire ring and let M be a torsion-free module. For each prime $\mathfrak{p}$ of A show that the natural map $M \to M_{\mathfrak{p}}$ is injective. In particular $A \to A_{\mathfrak{p}}$ is injective, but you can see that directly from the imbedding of A in its quotient field K.

Projective modules over Dedekind rings

For the next exercise we assume you have done the exercises on Dedekind rings in the preceding chapter. We shall see that for such rings, some parts of their module theory can be reduced to the case of principal rings by localization. We let $\mathfrak{o}$ be a Dedekind ring and K its quotient field.

11. Let M be a finitely generated torsion-free module over $\mathfrak{o}$. Prove that M is projective. [*Hint*: Given a prime ideal $\mathfrak{p}$, the localized module $M_{\mathfrak{p}}$ is finitely generated torsion-free over $\mathfrak{o}_{\mathfrak{p}}$, which is principal. Then $M_{\mathfrak{p}}$ is projective, so if F is finite free over $\mathfrak{o}$, and $f\colon F \to M$ is a surjective homomorphism, then $f_{\mathfrak{p}}\colon F_{\mathfrak{p}} \to M_{\mathfrak{p}}$ has a splitting $g_{\mathfrak{p}}\colon M_{\mathfrak{p}} \to F_{\mathfrak{p}}$, such that $f_{\mathfrak{p}} \circ g_{\mathfrak{p}} = \mathrm{id}_{M_{\mathfrak{p}}}$. There exists $c_{\mathfrak{p}} \in \mathfrak{o}$ such that $c_{\mathfrak{p}} \notin \mathfrak{p}$ and $c_{\mathfrak{p}} g_{\mathfrak{p}}(M) \subset F$. The family $\{c_{\mathfrak{p}}\}$ generates the unit ideal $\mathfrak{o}$ (why?), so there is a finite number of elements $c_{\mathfrak{p}_i}$ and elements $x_i \in \mathfrak{o}$ such that $\sum x_i c_{\mathfrak{p}_i} = 1$. Let

$$g = \sum x_i c_{\mathfrak{p}_i} g_{\mathfrak{p}_i}.$$

Then show that $g\colon M \to F$ gives a homomorphism such that $f \circ g = \mathrm{id}_M$.]

12. (a) Let $\mathfrak{a}, \mathfrak{b}$ be ideals. Show that there is an isomorphism of $\mathfrak{o}$-modules

$$\mathfrak{a} \oplus \mathfrak{b} \xrightarrow{\approx} \mathfrak{o} \oplus \mathfrak{a}\mathfrak{b}$$

[*Hint*: First do this when $\mathfrak{a}, \mathfrak{b}$ are relatively prime. Consider the homomorphism $\mathfrak{a} \oplus \mathfrak{b} \to \mathfrak{a} + \mathfrak{b}$, and use Exercise 10. Reduce the general case to the relatively prime case by using Exercise 19 of Chapter II.]

(b) Let $\mathfrak{a}, \mathfrak{b}$ be fractional ideals, and let $f\colon \mathfrak{a} \to \mathfrak{b}$ be an isomorphism (of $\mathfrak{o}$-modules, of course). Then f has an extension to a K-linear map $f_K\colon K \to K$. Let $c = f_K(1)$. Show that $\mathfrak{b} = c\mathfrak{a}$ and that f is given by the mapping $m_c\colon x \to cx$ (multiplication by c).

(c) Let $\mathfrak{a}$ be a fractional ideal. For each $b \in \mathfrak{a}^{-1}$ the map $m_b\colon \mathfrak{a} \to \mathfrak{o}$ is an element of the dual $\mathfrak{a}^{\vee}$. Show that $\mathfrak{a}^{-1} = \mathfrak{a}^{\vee} = \mathrm{Hom}_{\mathfrak{o}}(\mathfrak{a}, \mathfrak{o})$ under this map, and so $\mathfrak{a}^{\vee\vee} = \mathfrak{a}$.

13. (a) Let M be a projective finite module over the Dedekind ring $\mathfrak{o}$. Show that there exist free modules F and F' such that $F \supset M \supset F'$, and F, F' have the same rank, which is called the **rank** of M.

(b) Prove that there exists a basis $\{e_1, \ldots, e_n\}$ of F and ideals $\mathfrak{a}_1, \ldots, \mathfrak{a}_n$ such that $M = \mathfrak{a}_1 e_1 + \cdots + \mathfrak{a}_n e_n$, or in other words, $M \approx \bigoplus \mathfrak{a}_i$.

(c) Prove that $M \approx \mathfrak{o}^{n-1} \oplus \mathfrak{a}$ for some ideal $\mathfrak{a}$, and that the association $M \mapsto \mathfrak{a}$ induces an isomorphism of $K_0(\mathfrak{o})$ with the group of ideal classes $\text{Pic}(\mathfrak{o})$. (The group $K_0(\mathfrak{o})$ is the group of equivalence classes of projective modules defined at the end of §4.)

A few snakes

14. Consider a commutative diagram of R-modules and homomorphisms such that each row is exact:

$$\begin{array}{ccccccc} & & M' & \longrightarrow & M & \longrightarrow & M'' & \longrightarrow & 0 \\ & & \downarrow f & & \downarrow g & & \downarrow h & & \\ 0 & \longrightarrow & N' & \longrightarrow & N & \longrightarrow & N'' & & \end{array}$$

Prove:

(a) If f, h are monomorphisms then g is a monomorphism.

(b) If f, h are surjective, then g is surjective.

(c) Assume in addition that $0 \to M' \to M$ is exact and that $N \to N'' \to 0$ is exact. Prove that if any two of f, g, h are isomorphisms, then so is the third. [*Hint*: Use the snake lemma.]

15. **The five lemma.** Consider a commutative diagram of R-modules and homomorphisms such that each row is exact:

$$\begin{array}{ccccccccc} M_1 & \longrightarrow & M_2 & \longrightarrow & M_3 & \longrightarrow & M_4 & \longrightarrow & M_5 \\ \downarrow f_1 & & \downarrow f_2 & & \downarrow f_3 & & \downarrow f_4 & & \downarrow f_5 \\ N_1 & \longrightarrow & N_2 & \longrightarrow & N_3 & \longrightarrow & N_4 & \longrightarrow & N_5 \end{array}$$

Prove:

(a) If f_1 is surjective and f_2, f_4 are monomorphisms, then f_3 is a monomorphism.

(b) If f_5 is a monomorphism and f_2, f_4 are surjective, then f_3 is surjective. [*Hint*: Use the snake lemma.]

Inverse limits

16. Prove that the inverse limit of a system of simple groups in which the homomorphisms are surjective is either the trivial group, or a simple group.

17. (a) Let n range over the positive integers and let p be a prime number. Show that the abelian groups $A_n = \mathbf{Z}/p^n\mathbf{Z}$ form a projective system under the canonical homomorphism if $n \geqq m$. Let $\mathbf{Z}_p$ be its inverse limit. Show that $\mathbf{Z}_p$ maps surjectively on each $\mathbf{Z}/p^n\mathbf{Z}$; that $\mathbf{Z}_p$ has no divisors of 0, and has a unique maximal ideal generated by p. Show that $\mathbf{Z}_p$ is factorial, with only one prime, namely p itself.

(b) Next consider all ideals of $\mathbf{Z}$ as forming a directed system, by divisibility. Prove that

$$\varprojlim_{(a)} \mathbf{Z}/(a) = \prod_p \mathbf{Z}_p,$$

where the limit is taken over all ideals (a), and the product is taken over all primes p.

18. (a) Let $\{A_n\}$ be an inversely directed sequence of commutative rings, and let $\{M_n\}$ be an inversely directed sequence of modules, M_n being a module over A_n such that the following diagram is commutative:

$$\begin{array}{ccccc} A_{n+1} & \times M_{n+1} & \to M_{n+1} \\ \downarrow & \downarrow & \downarrow \\ A_n & \times M_n & \to M_n \end{array}$$

The vertical maps are the homomorphisms of the directed sequence, and the horizontal maps give the operation of the ring on the module. Show that $\varprojlim M_n$ is a module over $\varprojlim A_n$.

(b) Let M be a p-divisible group. Show that $T_p(A)$ is a module over $\mathbf{Z}_p$.

(c) Let M, N be p-divisible groups. Show that $T_p(M \oplus N) = T_p(M) \oplus T_p(N)$, as modules over $\mathbf{Z}_p$.

Direct limits

19. Let (A_i, f_j^i) be a directed family of modules. Let $a_k \in A_k$ for some k, and suppose that the image of a_k in the direct limit A is 0. Show that there exists some index $j \geqq k$ such that $f_j^k(a_k) = 0$. In other words, whether some element in some group A_i vanishes in the direct limit can already be seen within the original data. One way to see this is to use the construction of Theorem 10.1.

20. Let I, J be two directed sets, and give the product $I \times J$ the obvious ordering that $(i, j) \leqq (i', j')$ if $i \leqq i'$ and $j \leqq j'$. Let A_{ij} be a family of abelian groups, with homomorphisms indexed by $I \times J$, and forming a directed family. Show that the direct limits

$$\varinjlim_i \varinjlim_j A_{ij} \quad \text{and} \quad \varinjlim_j \varinjlim_i A_{ij}$$

exist and are isomorphic in a natural way. State and prove the same result for inverse limits.

21. Let (M'_i, f_j^i), (M_i, g_j^i) be directed systems of modules over a ring. By a **homomorphism**

$$(M'_i) \xrightarrow{u} (M_i)$$

one means a family of homomorphisms $u_i : M'_i \to M_i$ for each i which commute with the f_j^i, g_j^i. Suppose we are given an exact sequence

$$0 \to (M'_i) \xrightarrow{u} (M_i) \xrightarrow{v} (M''_i) \to 0$$

of directed systems, meaning that for each i, the sequence

$$0 \to M'_i \to M_i \to M''_i \to 0$$

is exact. Show that the direct limit preserves exactness, that is

$$0 \to \varinjlim M_i' \to \varinjlim M_i \to \varinjlim M_i'' \to 0$$

is exact.

22. (a) Let $\{M_i\}$ be a family of modules over a ring. For any module N show that

$$\mathrm{Hom}(\bigoplus M_i, N) = \prod \mathrm{Hom}(M_i, N)$$

(b) Show that

$$\mathrm{Hom}(N, \prod M_i) = \prod \mathrm{Hom}(N, M_i).$$

23. Let $\{M_i\}$ be a directed family of modules over a ring. For any module N show that

$$\varinjlim \mathrm{Hom}(N, M_i) = \mathrm{Hom}(N, \varinjlim M_i)$$

24. Show that any module is a direct limit of finitely generated submodules.

A module M is called **finitely presented** if there is an exact sequence

$$F_1 \to F_0 \to M \to 0$$

where F_0, F_1 are free with finite bases. The image of F_1 in F_0 is said to be the submodule of **relations**, among the free basis elements of F_0.

25. Show that any module is a direct limit of finitely presented modules (not necessarily submodules). In other words, given M, there exists a directed system $\{M_i, f_j^i\}$ with M_i finitely presented for all i such that

$$M \approx \varinjlim M_i.$$

[*Hint*: Any finitely generated submodule is such a direct limit, since an infinitely generated module of relations can be viewed as a limit of finitely generated modules of relations. Make this precise to get a proof.]

26. Let E be a module over a ring. Let $\{M_i\}$ be a directed family of modules. If E is finitely generated, show that the natural homomorphism

$$\varinjlim \mathrm{Hom}(E, M_i) \to \mathrm{Hom}(E, \varinjlim M_i)$$

is injective. If E is finitely presented, show that this homomorphism is an isomorphism. *Hint*: First prove the statements when E is free with finite basis. Then, say E is finitely presented by an exact sequence $F_1 \to F_0 \to E \to 0$. Consider the diagram:

$$\begin{array}{ccccccc}
0 & \longrightarrow & \varinjlim \mathrm{Hom}(E, M_i) & \longrightarrow & \varinjlim \mathrm{Hom}(F_0, M_i) & \longrightarrow & \varinjlim \mathrm{Hom}(F_1, M_i) \\
 & & \downarrow & & \downarrow & & \downarrow \\
0 & \longrightarrow & \mathrm{Hom}(E, \varinjlim M_i) & \longrightarrow & \mathrm{Hom}(F_0, \varinjlim M_i) & \longrightarrow & \mathrm{Hom}(F_1, \varinjlim M_i)
\end{array}$$

Graded Algebras

Let A be an algebra over a field k. By a **filtration** of A we mean a sequence of k-vector spaces A_i $(i = 0, 1, \ldots)$ such that

$$A_0 \subset A_1 \subset A_2 \subset \cdots \qquad \text{and} \qquad \bigcup A_i = A,$$

and $A_i A_j \subset A_{i+j}$ for all $i, j \geqq 0$. In particular, A is an A_0-algebra. We then call A a filtered algebra. Let R be an algebra. We say that R is **graded** if R is a direct sum $R = \bigoplus R_i$ of subspaces such that $R_i R_j \subset R_{i+j}$ for all $i, j \geqq 0$.

27. Let A be a filtered algebra. Define R_i for $i \geqq 0$ by $R_i = A_i/A_{i-1}$. By definition, $A_{-1} = \{0\}$. Let $R = \bigoplus R_i$, and $R_i = \mathrm{gr}_i(A)$. Define a natural product on R making R into a graded algebra, denoted by $\mathrm{gr}(A)$, and called the **associated graded algebra**.

28. Let A, B be filtered algebras, $A = \bigcup A_i$ and $B = \bigcup B_i$. Let $L\colon A \to B$ be an (A_0, B_0)-linear map preserving the filtration, that is $L(A_i) \subset B_i$ for all i, and $L(ca) = L(c)L(a)$ for $c \in A_0$ and $a \in A_i$ for all i.
    (a) Show that L induces an (A_0, B_0)-linear map

    $$\mathrm{gr}_i(L)\colon \mathrm{gr}_i(A) \to \mathrm{gr}_i(B) \qquad \text{for all } i.$$

    (b) Suppose that $\mathrm{gr}_i(L)$ is an isomorphism for all i. Show that L is an (A_0, B_0)-isomorphism.

29. Suppose k has characteristic 0. Let $\mathfrak{n}$ be the set of all strictly upper triangular matrices of a given size $n \times n$ over k.
    (a) For a given matrix $X \in \mathfrak{n}$, let $D_1(X), \ldots, D_n(X)$ be its diagonals, so $D_1 = D_1(X)$ is the main diagonal, and is 0 by the definition of $\mathfrak{n}$. Let $\mathfrak{n}_i$ be the subset of $\mathfrak{n}$ consisting of those matrices whose diagonals $D_1, \ldots, D_{n-i}$ are 0. Thus $\mathfrak{n}_0 = \{0\}$, $\mathfrak{n}_1$ consists of all matrices whose components are 0 except possibly for x_{nn}; $\mathfrak{n}_2$ consists of all matrices whose components are 0 except possibly those in the last two diagonals; and so forth. Show that each $\mathfrak{n}_i$ is an algebra, and its elements are nilpotent (in fact the $(i+1)$-th power of its elements is 0).
    (b) Let U be the set of elements $I + X$ with $X \in \mathfrak{n}$. Show that U is a multiplicative group.
    (c) Let exp be the exponential series defined as usual. Show that exp defines a polynomial function on $\mathfrak{n}$ (all but a finite number of terms are 0 when evaluated on a nilpotent matrix), and establishes a bijection

    $$\exp\colon \mathfrak{n} \to U.$$

    Show that the inverse is given by the standard log series.

CHAPTER IV

Polynomials

This chapter provides a continuation of Chapter II, §3. We prove standard properties of polynomials. Most readers will be acquainted with some of these properties, especially at the beginning for polynomials in one variable. However, one of our purposes is to show that some of these properties also hold over a commutative ring when properly formulated. The Gauss lemma and the reduction criterion for irreducibility will show the importance of working over rings. Chapter IX will give examples of the importance of working over the integers **Z** themselves to get universal relations. It happens that certain statements of algebra are universally true. To prove them, one proves them first for elements of a polynomial ring over **Z**, and then one obtains the statement in arbitrary fields (or commutative rings as the case may be) by specialization. The Cayley–Hamilton theorem of Chapter XV, for instance, can be proved in that way.

The last section on power series shows that the basic properties of polynomial rings can be formulated so as to hold for power series rings. I conclude this section with several examples showing the importance of power series in various parts of mathematics.

§1. BASIC PROPERTIES FOR POLYNOMIALS IN ONE VARIABLE

We start with the Euclidean algorithm.

Theorem 1.1. *Let A be a commutative ring, let f, $g \in A[X]$ be polynomials in one variable, of degrees $\geqq 0$, and assume that the leading*

coefficient of g *is a unit in* A. *Then there exist unique polynomials* $q, r \in A[X]$ *such that*

$$f = gq + r$$

and $\deg r < \deg g$.

Proof. Write

$$f(X) = a_n X^n + \cdots + a_0,$$
$$g(X) = b_d X^d + \cdots + b_0,$$

where $n = \deg f$, $d = \deg g$ so that a_n, $b_d \neq 0$ and b_d is a unit in A. We use induction on n.

If $n = 0$, and $\deg g > \deg f$, we let $q = 0$, $r = f$. If $\deg g = \deg f = 0$, then we let $r = 0$ and $q = a_n b_d^{-1}$.

Assume the theorem proved for polynomials of degree $< n$ (with $n > 0$). We may assume $\deg g \leqq \deg f$ (otherwise, take $q = 0$ and $r = f$). Then

$$f(X) = a_n b_d^{-1} X^{n-d} g(X) + f_1(X),$$

where $f_1(X)$ has degree $< n$. By induction, we can find q_1, r such that

$$f(X) = a_n b_d^{-1} X^{n-d} g(X) + q_1(X) g(X) + r(X)$$

and $\deg r < \deg g$. Then we let

$$q(X) = a_n b_d^{-1} X^{n-d} + q_1(X)$$

to conclude the proof of existence for q, r.

As for uniqueness, suppose that

$$f = q_1 g + r_1 = q_2 g + r_2$$

with $\deg r_1 < \deg g$ and $\deg r_2 < \deg g$. Subtracting yields

$$(q_1 - q_2) g = r_2 - r_1.$$

Since the leading coefficient of g is assumed to be a unit, we have

$$\deg(q_1 - q_2) g = \deg(q_1 - q_2) + \deg g.$$

Since $\deg(r_2 - r_1) < \deg g$, this relation can hold only if $q_1 - q_2 = 0$, i.e. $q_1 = q_2$, and hence finally $r_1 = r_2$ as was to be shown.

Theorem 1.2. *Let* k *be a field. Then the polynomial ring in one variable* $k[X]$ *is principal.*

Proof. Let $\mathfrak{a}$ be an ideal of $k[X]$, and assume $\mathfrak{a} \neq 0$. Let g be an element of $\mathfrak{a}$ of smallest degree $\geqq 0$. Let f be any element of $\mathfrak{a}$ such that $f \neq 0$. By the Euclidean algorithm we can find $q, r \in k[X]$ such that

$$f = qg + r$$

and $\deg r < \deg g$. But $r = f - qg$, whence r is in $\mathfrak{a}$. Since g had minimal degree $\geqq 0$ it follows that $r = 0$, hence that $\mathfrak{a}$ consists of all polynomials qg (with $q \in k[X]$). This proves our theorem. By Theorem 5.2 of Chapter II we get:

Corollary 1.3. *The ring $k[X]$ is factorial.*

If k is a field then every non-zero element of k is a unit in k, and one sees immediately that the units of $k[X]$ are simply the units of k. (No polynomial of degree $\geqq 1$ can be a unit because of the addition formula for the degree of a product.)

A polynomial $f(X) \in k[X]$ is called **irreducible** if it has degree $\geqq 1$, and if one cannot write $f(X)$ as a product

$$f(X) = g(X)h(X)$$

with $g, h \in k[X]$, and both $g, h \notin k$. Elements of k are usually called **constant polynomials**, so we can also say that in such a factorization, one of g or h must be constant. A polynomial is called **monic** if it has leading coefficient 1.

Let A be a commutative ring and $f(X)$ a polynomial in $A[X]$. Let A be a subring of B. An element $b \in B$ is called a **root** or a **zero** of f in B if $f(b) = 0$. Similarly, if (X) is an n-tuple of variables, an n-tuple (b) is called a zero of f if $f(b) = 0$.

Theorem 1.4. *Let k be a field and f a polynomial in one variable X in $k[X]$, of degree $n \geqq 0$. Then f has at most n roots in k, and if a is a root of f in k, then $X - a$ divides $f(X)$.*

Proof. Suppose $f(a) = 0$. Find q, r such that

$$f(X) = q(X)(X - a) + r(X)$$

and $\deg r < 1$. Then

$$0 = f(a) = r(a).$$

Since $r = 0$ or r is a non-zero constant, we must have $r = 0$, whence $X - a$ divides $f(X)$. If $a_1, \ldots, a_m$ are distinct roots of f in k, then inductively we see that the product

$$(X - a_1) \cdots (X - a_m)$$

divides $f(X)$, whence $m \leqq n$, thereby proving the theorem. The next corollaries give applications of Theorem 1.4 to polynomial functions.

Corollary 1.5. *Let k be a field and T an infinite subset of k. Let $f(X) \in k[X]$ be a polynomial in one variable. If $f(a) = 0$ for all $a \in T$, then $f = 0$, i.e. f induces the zero function.*

Corollary 1.6. *Let k be a field, and let $S_1, \ldots, S_n$ be infinite subsets of k. Let $f(X_1, \ldots, X_n)$ be a polynomial in n variables over k. If $f(a_1, \ldots, a_n) = 0$ for all $a_i \in S_i$ $(i = 1, \ldots, n)$, then $f = 0$.*

Proof. By induction. We have just seen the result is true for one variable. Let $n \geqq 2$, and write

$$f(X_1, \ldots, X_n) = \sum_j f_i(X_1, \ldots, X_{n-1})X_n^j$$

as a polynomial in X_n with coefficients in $k[X_1, \ldots, X_{n-1}]$. If there exists

$$(b_1, \ldots, b_{n-1}) \in S_1 \times \cdots \times S_{n-1}$$

such that for some j we have $f_i(b_1, \ldots, b_{n-1}) \neq 0$, then

$$f(b_1, \ldots, b_{n-1}, X_n)$$

is a non-zero polynomial in $k[X_n]$ which takes on the value 0 for the infinite set of elements S_n. This is impossible. Hence f_j induces the zero function on $S_1 \times \cdots \times S_{n-1}$ for all j, and by induction we have $f_j = 0$ for all j. Hence $f = 0$, as was to be shown.

Corollary 1.7. *Let k be an infinite field and f a polynomial in n variables over k. If f induces the zero function on $k^{(n)}$, then $f = 0$.*

We shall now consider the case of finite fields. Let k be a finite field with q elements. Let $f(X_1, \ldots, X_n)$ be a polynomial in n variables over k. Write

$$f(X_1, \ldots, X_n) = \sum a_{(\nu)} X_1^{\nu_1} \cdots X_n^{\nu_n}.$$

If $a_{(\nu)} \neq 0$, we recall that the monomial $M_{(\nu)}(X)$ **occurs** in f. Suppose this is the case, and that in this monomial $M_{(\nu)}(X)$, some variable X_i occurs with an exponent $\nu_i \geqq q$. We can write

$$X_i^{\nu_i} = X_i^{q+\mu}, \qquad \mu = \text{integer} \geqq 0.$$

If we now replace $X_i^{\nu_i}$ by $X_i^{\mu+1}$ in this monomial, then we obtain a new polynomial which gives rise to the same function as f. The degree of this new polynomial is at most equal to the degree of f.

Performing the above operation a finite number of times, for all the monomials occurring in f and all the variables $X_1, \ldots, X_n$ we obtain some polynomial f^* giving rise to the same function as f, but whose degree in each variable is $< q$.

Corollary 1.8. *Let k be a finite field with q elements. Let f be a polynomial in n variables over k such that the degree of f in each variable is $<q$. If f induces the zero function on $k^{(n)}$, then $f = 0$.*

Proof. By induction. If $n = 1$, then the degree of f is $< q$, and hence f cannot have q roots unless it is 0. The inductive step is carried out just as we did for the proof of Corollary 1.6 above.

Let f be a polynomial in n variables over the finite field k. A polynomial g whose degree in each variable is $< q$ will be said to be **reduced**. We have shown above that there exists a reduced polynomial f^* which gives the same function as f on $k^{(n)}$. Theorem 1.8 now shows that *this reduced polynomial is unique.* Indeed, if g_1, g_2 are reduced polynomials giving the same function, then $g_1 - g_2$ is reduced and gives the zero function. Hence $g_1 - g_2 = 0$ and $g_1 = g_2$.

We shall give one more application of Theorem 1.4. Let k be a field. By a **multiplicative subgroup** of k we shall mean a subgroup of the group k^* (non-zero elements of k).

Theorem 1.9. *Let k be a field and let U be a finite multiplicative subgroup of k. Then U is cyclic.*

Proof. Write U as a product of subgroups $U(p)$ for each prime p, where $U(p)$ is a p-group. By Proposition 4.3(vi) of Chapter I, it will suffice to prove that $U(p)$ is cyclic for each p. Let a be an element of $U(p)$ of maximal period p^r for some integer r. Then $x^{p^r} = 1$ for every element $x \in U(p)$, and hence all elements of $U(p)$ are roots of the polynomial

$$X^{p^r} - 1.$$

The cyclic group generated by a has p^r elements. If this cyclic group is not equal to $U(p)$, then our polynomial has more than p^r roots, which is impossible. Hence a generates $U(p)$, and our theorem is proved.

Corollary 1.10. *If k is a finite field, then k^* is cyclic.*

An element ζ in a field k such that there exists an integer $n \geqq 1$ such that $\zeta^n = 1$ is called a **root of unity**, or more precisely an n-th root of unity. Thus the set of n-th roots of unity is the set of roots of the polynomial $X^n - 1$. There are at most n such roots, and they obviously form a group, which is

cyclic by Theorem 1.9. We shall study roots of unity in greater detail later. A generator for the group of n-th roots of unity is called a **primitive** n-th root of unity. For example, in the complex numbers, $e^{2\pi i/n}$ is a primitive n-th root of unity, and the n-th roots of unity are of type $e^{2\pi i\nu/n}$ with $1 \leqq \nu \leqq n$.

The group of roots of unity is denoted by $\boldsymbol{\mu}$. The group of roots of unity in a field K is denoted by $\boldsymbol{\mu}(K)$.

A field k is said to be **algebraically closed** if every polynomial in $k[X]$ of degree $\geqq 1$ has a root in k. In books on analysis, it is proved that the complex numbers are algebraically closed. In Chapter V we shall prove that a field k is always contained in some algebraically closed field. If k is algebraically closed then the irreducible polynomials in $k[X]$ are the polynomials of degree 1. In such a case, the unique factorization of a polynomial f of degree $\geqq 0$ can be written in the form

$$f(X) = c \prod_{i=1}^{r} (X - \alpha_i)^{m_i}$$

with $c \in k$, $c \neq 0$ and distinct roots $\alpha_1, \ldots, \alpha_r$. We next develop a test when $m_i > 1$.

Let A be a commutative ring. We define a map

$$D: A[X] \to A[X]$$

of the polynomial ring into itself. If $f(X) = a_n X^n + \cdots + a_0$ with $a_i \in A$, we define the **derivative**

$$Df(X) = f'(X) = \sum_{\nu=1}^{n} \nu a_\nu X^{\nu-1} = n a_n X^{n-1} + \cdots + a_1.$$

One verifies easily that if f, g are polynomials in $A[X]$, then

$$(f + g)' = f' + g', \qquad (fg)' = f'g + fg',$$

and if $a \in A$, then

$$(af)' = af'.$$

Let K be a field and f a non-zero polynomial in $K[X]$. Let a be a root of f in K. We can write

$$f(X) = (X - a)^m g(X)$$

with some polynomial $g(X)$ relatively prime to $X - a$ (and hence such that $g(a) \neq 0$). We call m the **multiplicity** of a in f, and say that a is a **multiple root** if $m > 1$.

Proposition 1.11. *Let K, f be as above. The element a of K is a multiple root of f if and only if it is a root and $f'(a) = 0$.*

Proof. Factoring f as above, we get

$$f'(X) = (X - a)^m g'(X) + m(X - a)^{m-1} g(X).$$

If $m > 1$, then obviously $f'(a) = 0$. Conversely, if $m = 1$ then

$$f'(X) = (X - a)g'(X) + g(X),$$

whence $f'(a) = g(a) \neq 0$. Hence if $f'(a) = 0$ we must have $m > 1$, as desired.

Proposition 1.12. *Let $f \in K[X]$. If K has characteristic 0, and f has degree $\geqq 1$, then $f' \neq 0$. Let K have characteristic $p > 0$ and f have degree $\geqq 1$. Then $f' = 0$ if and only if, in the expression for $f(X)$ given by*

$$f(X) = \sum_{\nu=1}^{n} a_\nu X^\nu,$$

p divides each integer ν such that $a_\nu \neq 0$.

Proof. If K has characteristic 0, then the derivative of a monomial $a_\nu X^\nu$ such that $\nu \geqq 1$ and $a_\nu \neq 0$ is not zero since it is $\nu a_\nu X^{\nu-1}$. If K has characteristic $p > 0$, then the derivative of such a monomial is 0 if and only if $p | \nu$, as contended.

Let K have characteristic $p > 0$, and let f be written as above, and be such that $f'(X) = 0$. Then one can write

$$f(X) = \sum_{\mu=1}^{d} b_\mu X^{p\mu}$$

with $b_\mu \in K$.

Since the binomial coefficients $\binom{p}{\nu}$ are divisible by p for $1 \leqq \nu \leqq p - 1$ we see that if K has characteristic p, then for $a, b \in K$ we have

$$(a + b)^p = a^p + b^p.$$

Since obviously $(ab)^p = a^p b^p$, the map

$$x \mapsto x^p$$

is a homomorphism of K into itself, which has trivial kernel, hence is injective. Iterating, we conclude that for each integer $r \geqq 1$, the map $x \mapsto x^{p^r}$

is an endomorphism of K, called the **Frobenius endomorphism**. Inductively, if $c_1, \ldots, c_n$ are elements of K, then

$$(c_1 + \cdots + c_n)^p = c_1^p + \cdots + c_n^p.$$

Applying these remarks to polynomials, we see that for any element $a \in K$ we have

$$(X - a)^{p^r} = X^{p^r} - a^{p^r}.$$

If $c \in K$ and the polynomial

$$X^{p^r} - c$$

has one root a in K, then $a^{p^r} = c$ and

$$X^{p^r} - c = (X - a)^{p^r}.$$

Hence our polynomial has precisely one root, of multiplicity p^r. For instance, $(X - 1)^{p^r} = X^{p^r} - 1$.

§2. POLYNOMIALS OVER A FACTORIAL RING

Let A be a factorial ring, and K its quotient field. Let $a \in K$, $a \neq 0$. We can write a as a quotient of elements in A, having no prime factor in common. If p is a prime element of A, then we can write

$$a = p^r b,$$

where $b \in K$, r is an integer, and p does not divide the numerator or denominator of b. Using the unique factorization in A, we see at once that r is uniquely determined by a, and we call r the **order of a at p** (and write $r = \text{ord}_p\, a$). If $a = 0$, we define its order at p to be ∞.

If $a, a' \in K$ and $aa' \neq 0$, then

$$\text{ord}_p(aa') = \text{ord}_p\, a + \text{ord}_p\, a'.$$

This is obvious.

Let $f(X) \in K[X]$ be a polynomial in one variable, written

$$f(X) = a_0 + a_1X + \cdots + a_nX^n.$$

If $f = 0$, we define $\text{ord}_p f$ to be ∞. If $f \neq 0$, we define $\text{ord}_p f$ to be

$$\text{ord}_p f = \min \text{ord}_p a_i,$$

the minimum being taken over all those i such that $a_i \neq 0$.

If $r = \text{ord}_p f$, we call up^r a **p-content** for f, if u is any unit of A. We define the **content** of f to be the product.

$$\prod p^{\text{ord}_p f},$$

the product being taken over all p such that $\text{ord}_p f \neq 0$, or any multiple of this product by a unit of A. Thus the content is well defined up to multiplication by a unit of A. We abbreviate **content** by **cont**.

If $b \in K$, $b \neq 0$, then $\text{cont}(bf) = b\,\text{cont}(f)$. This is clear. Hence we can write

$$f(X) = c \cdot f_1(X)$$

where $c = \text{cont}(f)$, and $f_1(X)$ has content 1. In particular, all coefficients of f_1 lie in A, and their g.c.d. is 1. We define a polynomial with content 1 to be a **primitive polynomial**.

Theorem 2.1. (Gauss Lemma). *Let A be a factorial ring, and let K be its quotient field. Let $f, g \in K[X]$ be polynomials in one variable. Then*

$$\text{cont}(fg) = \text{cont}(f)\,\text{cont}(g).$$

Proof. Writing $f = cf_1$ and $g = dg_1$ where $c = \text{cont}(f)$ and $d = \text{cont}(g)$, we see that it suffices to prove: If f, g have content 1, then fg also has content 1, and for this, it suffices to prove that for each prime p, $\text{ord}_p(fg) = 0$. Let

$$f(X) = a_n X^n + \cdots + a_0, \qquad a_n \neq 0,$$
$$g(X) = b_m X^m + \cdots + b_0, \qquad b_m \neq 0,$$

be polynomials of content 1. Let p be a prime of A. It will suffice to prove that p does not divide all coefficients of fg. Let r be the largest integer such that $0 \leqq r \leqq n$, $a_r \neq 0$, and p does not divide a_r. Similarly, let b_s be the coefficient of g farthest to the left, $b_s \neq 0$, such that p does not divide b_s. Consider the coefficient of X^{r+s} in $f(X)g(X)$. This coefficient is equal to

$$c = a_r b_s + a_{r+1} b_{s-1} + \cdots$$
$$+ a_{r-1} b_{s+1} + \cdots$$

and $p \nmid a_r b_s$. However, p divides every other non-zero term in this sum since in each term there will be some coefficient a_i to the left of a_r or some coefficient b_j to the left of b_s. Hence p does not divide c, and our lemma is proved.

We shall now give another proof for the key step in the above argument, namely the statement:

If $f, g \in A[X]$ are primitive (i.e. have content 1) then fg is primitive.

Proof. We have to prove that a given prime p does not divide all the coefficients of fg. Consider reduction mod p, namely the canonical homomorphism $A \to A/(p) = \bar{A}$. Denote the image of a polynomial by a bar, so $f \mapsto \bar{f}$ and $g \mapsto \bar{g}$ under the reduction homomorphism. Then

$$\overline{fg} = \bar{f}\bar{g}.$$

By hypothesis, $\bar{f} \neq 0$ and $\bar{g} \neq 0$. Since $\bar{A}$ is entire, it follows that $\overline{fg} \neq 0$, as was to be shown.

Corollary 2.2. *Let $f(X) \in A[X]$ have a factorization $f(X) = g(X)h(X)$ in $K[X]$. If $c_g = \text{cont}(g)$, $c_h = \text{cont}(h)$, and $g = c_g g_1$, $h = c_h h_1$, then*

$$f(X) = c_g c_h g_1(X) h_1(X),$$

and $c_g c_h$ is an element of A. In particular, if $f, g \in A[X]$ have content 1, then $h \in A[X]$ also.

Proof. The only thing to be proved is $c_g c_h \in A$. But

$$\text{cont}(f) = c_g c_h \,\text{cont}(g_1 h_1) = c_g c_h,$$

whence our assertion follows.

Theorem 2.3. *Let A be a factorial ring. Then the polynomial ring $A[X]$ in one variable is factorial. Its prime elements are the primes of A and polynomials in $A[X]$ which are irreducible in $K[X]$ and have content 1.*

Proof. Let $f \in A[X]$, $f \neq 0$. Using the unique factorization in $K[X]$ and the preceding corollary, we can find a factorization

$$f(X) = c \cdot p_1(X) \cdots p_r(X)$$

where $c \in A$, and $p_1, \ldots, p_r$ are polynomials in $A[X]$ which are irreducible in $K[X]$. Extracting their contents, we may assume without loss of generality that the content of p_i is 1 for each i. Then $c = \text{cont}(f)$ by the Gauss lemma. This gives us the existence of the factorization. It follows that each $p_i(X)$ is irreducible in $A[X]$. If we have another such factorization, say

$$f(X) = d \cdot q_1(X) \cdots q_s(X),$$

then from the unique factorization in $K[X]$ we conclude that $r = s$, and after a permutation of the factors we have

$$p_i = a_i q_i$$

with elements $a_i \in K$. Since both p_i, q_i are assumed to have content 1, it follows that a_i in fact lies in A and is a unit. This proves our theorem.

Corollary 2.4. *Let A be a factorial ring. Then the ring of polynomials in n variables $A[X_1, \dots, X_n]$ is factorial. Its units are precisely the units of A, and its prime elements are either primes of A or polynomials which are irreducible in $K[X]$ and have content* 1.

Proof. Induction.

In view of Theorem 2.3, when we deal with polynomials over a factorial ring and having content 1, it is not necessary to specify whether such polynomials are irreducible over A or over the quotient field K. The two notions are equivalent.

Remark 1. The polynomial ring $K[X_1, \dots, X_n]$ over a field K is not principal when $n \geqq 2$. For instance, the ideal generated by $X_1, \dots, X_n$ is not principal (trivial proof).

Remark 2. It is usually not too easy to decide when a given polynomial (say in one variable) is irreducible. For instance, the polynomial $X^4 + 4$ is *reducible* over the rational numbers, because

$$X^4 + 4 = (X^2 - 2X + 2)(X^2 + 2X + 2).$$

Later in this book we shall give a precise criterion when a polynomial $X^n - a$ is irreducible. Other criteria are given in the next section.

§3. CRITERIA FOR IRREDUCIBILITY

The first criterion is:

Theorem 3.1. (Eisenstein's Criterion). *Let A be a factorial ring. Let K be its quotient field. Let $f(X) = a_nX^n + \cdots + a_0$ be a polynomial of degree $n \geqq 1$ in $A[X]$. Let p be a prime of A, and assume:*

$$a_n \not\equiv 0 \pmod{p}, \qquad a_i \equiv 0 \pmod{p} \qquad \textit{for all} \quad i < n,$$
$$a_0 \not\equiv 0 \pmod{p^2}.$$

Then $f(X)$ is irreducible in $K[X]$.

Proof. Extracting a g.c.d. for the coefficients of f, we may assume without loss of generality that the content of f is 1. If there exists a factorization into factors of degree $\geqq 1$ in $K[X]$, then by the corollary of Gauss' lemma there exists a factorization in $A[X]$, say $f(X) = g(X)h(X)$,

$$g(X) = b_d X^d + \cdots + b_0,$$

$$h(X) = c_m X^m + \cdots + c_0,$$

with d, $m \geqq 1$ and $b_d c_m \neq 0$. Since $b_0 c_0 = a_0$ is divisible by p but not p^2, it follows that one of b_0, c_0 is not divisible by p, say b_0. Then $p|c_0$. Since $c_m b_d = a_n$ is not divisible by p, it follows that p does not divide c_m. Let c_r be the coefficient of h furthest to the right such that $c_r \not\equiv 0 \pmod{p}$. Then

$$a_r = b_0 c_r + b_1 c_{r-1} + \cdots .$$

Since $p \nmid b_0 c_r$ but p divides every other term in this sum, we conclude that $p \nmid a_r$, a contradiction which proves our theorem.

Example. Let a be a non-zero square-free integer $\neq \pm 1$. Then for any integer $n \geqq 1$, the polynomial $X^n - a$ is irreducible over $\mathbf{Q}$. The polynomials $3X^5 - 15$ and $2X^{10} - 21$ are irreducible over $\mathbf{Q}$.

There are some cases in which a polynomial does not satisfy Eisenstein's criterion, but a simple transform of it does.

Example. Let p be a prime number. Then the polynomial

$$f(X) = X^{p-1} + \cdots + 1$$

is irreducible over $\mathbf{Q}$.

Proof. It will suffice to prove that the polynomial $f(X + 1)$ is irreducible over $\mathbf{Q}$. We note that the binomial coefficients

$$\binom{p}{\nu} = \frac{p!}{\nu!(p-\nu)!}, \qquad 1 \leqq \nu \leqq p - 1,$$

are divisible by p (because the numerator is divisible by p and the denominator is not, and the coefficient is an integer). We have

$$f(X+1) = \frac{(X+1)^p - 1}{(X+1) - 1} = \frac{X^p + pX^{p-1} + \cdots + pX}{X}$$

from which one sees that $f(X + 1)$ satisfies Eisenstein's criterion.

Example. Let E be a field and t an element of some field containing E such that t is transcendental over E. Let K be the quotient field of $E[t]$.

For any integer $n \geqq 1$ the polynomial $X^n - t$ is irreducible in $K[X]$. This comes from the fact that the ring $A = E[t]$ is factorial and that t is a prime in it.

Theorem 3.2. (Reduction Criterion). *Let A, B be entire rings, and let*

$$\varphi: A \to B$$

be a homomorphism. Let K, L be the quotient fields of A and B respectively. Let $f \in A[X]$ be such that $\varphi f \neq 0$ and $\deg \varphi f = \deg f$. If φf is irreducible in $L[X]$, then f does not have a factorization $f(X) = g(X)h(X)$ with

$$g, h \in A[X] \quad \text{and} \quad \deg g, \deg h \geqq 1.$$

Proof. Suppose f has such a factorization. Then $\varphi f = (\varphi g)(\varphi h)$. Since $\deg \varphi g \leqq \deg g$ and $\deg \varphi h \leqq \deg h$, our hypothesis implies that we must have equality in these degree relations. Hence from the irreducibility in $L[X]$ we conclude that g or h is an element of A, as desired.

In the preceding criterion, suppose that A is a local ring, i.e. a ring having a unique maximal ideal $\mathfrak{p}$, and that $\mathfrak{p}$ is the kernel of φ. Then from the irreducibility of φf in $L[X]$ we conclude the irreducibility of f in $A[X]$. Indeed, any element of A which does not lie in $\mathfrak{p}$ must be a unit in A, so our last conclusion in the proof can be strengthened to the statement that g or h is a unit in A.

One can also apply the criterion when A is factorial, and in that case deduce the irreducibility of f in $K[X]$.

Example. Let p be a prime number. It will be shown later that $X^p - X - 1$ is irreducible over the field $\mathbf{Z}/p\mathbf{Z}$. Hence $X^p - X - 1$ is irreducible over $\mathbf{Q}$. Similarly,

$$X^5 - 5X^4 - 6X - 1$$

is irreducible over $\mathbf{Q}$.

There is also a routine elementary school test whether a polynomial has a root or not.

Proposition 3.3. (Integral Root Test). *Let A be a factorial ring and K its quotient field. Let*

$$f(X) = a_n X^n + \cdots + a_0 \in A[X].$$

Let $\alpha \in K$ be a root of f, with $\alpha = b/d$ expressed with $b, d \in A$ and b, d relatively prime. Then $b \mid a_0$ and $d \mid a_n$. In particular, if the leading coefficient a_n is 1, then a root α must lie in A and divides a_0.

We leave the proof to the reader, who should be used to this one from way back. As an irreducibility test, the test is useful especially for a polynomial of degree 2 or 3, when reducibility is equivalent with the existence of a root in the given field.

§4. HILBERT'S THEOREM

This section proves a basic theorem of Hilbert concerning the ideals of a polynomial ring. We define a commutative ring A to be **Noetherian** if every ideal is finitely generated.

Theorem 4.1. *Let A be a commutative Noetherian ring. Then the polynomial ring $A[X]$ is also Noetherian.*

Proof. Let $\mathfrak{A}$ be an ideal of $A[X]$. Let a_i consist of 0 and the set of elements $a \in A$ appearing as leading coefficient in some polynomial

$$a_0 + a_1 X + \cdots + aX^i$$

lying in $\mathfrak{A}$. Then it is clear that $\mathfrak{a}_i$ is an ideal. (If a, b are in $\mathfrak{a}_i$, then $a \pm b$ is in $\mathfrak{a}_i$ as one sees by taking the sum and difference of the corresponding polynomials. If $x \in A$, then $xa \in \mathfrak{a}_i$ as one sees by multiplying the corresponding polynomial by x.) Furthermore we have

$$\mathfrak{a}_0 \subset \mathfrak{a}_1 \subset \mathfrak{a}_2 \subset \cdots,$$

in other words, our sequence of ideals $\{\mathfrak{a}_i\}$ is increasing. Indeed, to see this multiply the above polynomial by X to see that $a \in \mathfrak{a}_{i+1}$.

By criterion (2) of Chapter X, §1, the sequence of ideals $\{\mathfrak{a}_i\}$ stops, say at $\mathfrak{a}_r$:

$$\mathfrak{a}_0 \subset \mathfrak{a}_1 \subset \mathfrak{a}_2 \subset \cdots \subset \mathfrak{a}_r = \mathfrak{a}_{r+1} = \cdots.$$

Let

$$a_{01}, \ldots, a_{0n_0} \quad \text{be generators for} \quad \mathfrak{a}_0,$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

$$a_{r1}, \ldots, a_{rn_r} \quad \text{be generators for} \quad \mathfrak{a}_r.$$

For each $i = 0, \ldots, r$ and $j = 1, \ldots, n_i$ let f_{ij} be a polynomial in $\mathfrak{A}$, of degree i, with leading coefficient a_{ij}. We contend that the polynomials f_{ij} are a set of generators for $\mathfrak{A}$.

Let f be a polynomial of degree d in $\mathfrak{A}$. We shall prove that f is in the ideal generated by the f_{ij}, by induction on d. Say $d \geqq 0$. If $d > r$, then we

note that the leading coefficients of

$$X^{d-r}f_{r1}, \ldots, X^{d-r}f_{rn_r}$$

generate $\mathfrak{a}_d$. Hence there exist elements $c_1, \ldots, c_{n_r} \in A$ such that the polynomial

$$f - c_1 X^{d-r} f_{r1} - \cdots - c_{n_r} X^{d-r} f_{rn_r}$$

has degree $< d$, and this polynomial also lies in $\mathfrak{A}$. If $d \leqq r$, we can subtract a linear combination

$$f - c_1 f_{d1} - \cdots - c_{n_d} f_{dn_d}$$

to get a polynomial of degree $< d$, also lying in $\mathfrak{A}$. We note that the polynomial we have subtracted from f lies in the ideal generated by the f_{ij}. By induction, we can subtract a polynomial g in the ideal generated by the f_{ij} such that $f - g = 0$, thereby proving our theorem.

We note that if $\varphi: A \to B$ is a surjective homomorphism of commutative rings and A is Noetherian, so is B. Indeed, let $\mathfrak{b}$ be an ideal of B, so $\varphi^{-1}(\mathfrak{b})$ is an ideal of A. Then there is a finite number of generators $(a_1, \ldots, a_n)$ for $\varphi^{-1}(\mathfrak{b})$, and it follows since φ is surjective that $\mathfrak{b} = \varphi(\varphi^{-1}(\mathfrak{b}))$ is generated by $\varphi(a_1), \ldots, \varphi(a_n)$, as desired. As an application, we obtain:

Corollary 4.2. *Let A be a Noetherian commutative ring, and let $B = A[x_1, \ldots, x_m]$ be a commutative ring finitely generated over A. Then B is Noetherian.*

Proof. Use Theorem 4.1 and the preceding remark, representing B as a factor ring of a polynomial ring.

Ideals in polynomial rings will be studied more deeply in Chapter IX. The theory of Noetherian rings and modules will be developed in Chapter X.

§5. PARTIAL FRACTIONS

In this section, we analyze the quotient field of a principal ring, using the factoriality of the ring.

Theorem 5.1. *Let A be a principal entire ring, and let P be a set of representatives for its irreducible elements. Let K be the quotient field of A, and let $\alpha \in K$. For each $p \in P$ there exists an element $\alpha_p \in A$ and an integer $j(p) \geqq 0$, such that $j(p) = 0$ for almost all $p \in P$, α_p and $p^{j(p)}$ are*

relatively prime, and

$$\alpha = \sum_{p \in P} \frac{\alpha_p}{p^{j(p)}}.$$

If we have another such expression

$$\alpha = \sum_{p \in P} \frac{\beta_p}{p^{i(p)}},$$

then $j(p) = i(p)$ *for all* p, *and* $\alpha_p \equiv \beta_p \bmod p^{j(p)}$ *for all* p.

Proof. We first prove existence, in a special case. Let a, b be relatively prime non-zero elements of A. Then there exists $x, y \in A$ such that $xa + yb = 1$. Hence

$$\frac{1}{ab} = \frac{x}{b} + \frac{y}{a}.$$

Hence any fraction c/ab with $c \in A$ can be decomposed into a sum of two fractions (namely cx/b and cy/a) whose denominators divide b and a respectively. By induction, it now follows that any $\alpha \in K$ has an expression as stated in the theorem, except possibly for the fact that p may divide α_p. Canceling the greatest common divisor yields an expression satisfying all the desired conditions.

As for uniqueness, suppose that α has two expressions as stated in the theorem. Let q be a fixed prime in P. Then

$$\frac{\alpha_q}{q^{j(q)}} - \frac{\beta_q}{q^{i(q)}} = \sum_{p \neq q} \frac{\beta_p}{p^{i(p)}} - \frac{\alpha_p}{p^{j(p)}}.$$

If $j(q) = i(q) = 0$, our conditions concerning q are satisfied. Suppose one of $j(q)$ or $i(q) > 0$, say $j(q)$, and say $j(q) \geqq i(q)$. Let d be a least common multiple for all powers $p^{j(p)}$ and $p^{i(p)}$ such that $p \neq q$. Multiply the above equation by $dq^{j(q)}$. We get

$$d(\alpha_q - q^{j(q)-i(q)}\beta_q) = q^{j(q)}\beta$$

for some $\beta \in A$. Furthermore, q does not divide d. If $i(q) < j(q)$ then q divides α_q, which is impossible. Hence $i(q) = j(q)$. We now see that $q^{j(q)}$ divides $\alpha_q - \beta_q$, thereby proving the theorem.

We apply Theorem 5.1 to the polynomial ring $k[X]$ over a field k. We let P be the set of irreducible polynomials, normalized so as to have leading coefficient equal to 1. Then P is a set of representatives for all the irreducible elements of $k[X]$. In the expression given for α in Theorem 5.1, we can now divide α_p by $p^{j(p)}$, i.e. use the Euclidean algorithm, if $\deg \alpha_p \geqq \deg p^{j(p)}$. We denote the quotient field of $k[X]$ by $k(X)$, and call its elements **rational functions.**

Theorem 5.2. *Let $A = k[X]$ be the polynomial ring in one variable over a field k. Let P be the set of irreducible polynomials in $k[X]$ with leading coefficient* 1. *Then any element f of $k(X)$ has a unique expression*

$$f(X) = \sum_{p \in P} \frac{f_p(X)}{p(X)^{j(p)}} + g(X),$$

where f_p, g are polynomials, $f_p = 0$ if $j(p) = 0$, f_p is relatively prime to p if $j(p) > 0$, and $\deg f_p < \deg p^{j(p)}$ *if $j(p) > 0$.*

Proof. The existence follows at once from our previous remarks. The uniqueness follows from the fact that if we have two expressions, with elements f_p and φ_p respectively, and polynomials g, h, then $p^{j(p)}$ divides $f_p - \varphi_p$, whence $f_p - \varphi_p = 0$, and therefore $f_p = \varphi_p$, $g = h$.

One can further decompose the term $f_p/p^{j(p)}$ by expanding f_p according to powers of p. One can in fact do something more general.

Theorem 5.3. *Let k be a field and $k[X]$ the polynomial ring in one variable. Let f, $g \in k[X]$, and assume* $\deg g \geqq 1$. *Then there exist unique polynomials*

$$f_0, f_1, \ldots, f_d \in k[X]$$

such that $\deg f_i < \deg g$ *and such that*

$$f = f_0 + f_1 g + \cdots + f_d g^d.$$

Proof. We first prove existence. If $\deg g > \deg f$, then we take $f_0 = f$ and $f_i = 0$ for $i > 0$. Suppose $\deg g \leqq \deg f$. We can find polynomials q, r with $\deg r < \deg g$ such that

$$f = qg + r,$$

and since $\deg g \geqq 1$ we have $\deg q < \deg f$. Inductively, there exist polynomials $h_0, h_1, \ldots, h_s$ such that

$$q = h_0 + h_1 g + \cdots + h_s g^s,$$

and hence

$$f = r + h_0 g + \cdots + h_s g^{s+1},$$

thereby proving existence.

As for uniqueness, let

$$f = f_0 + f_1 g + \cdots + f_d g^d = \varphi_0 + \varphi_1 g + \cdots + \varphi_m g^m$$

be two expressions satisfying the conditions of the theorem. Adding terms

equal to 0 to either side, we may assume that $m = d$. Subtracting, we get

$$0 = (f_0 - \varphi_0) + \cdots + (f_d - \varphi_d)g^d.$$

Hence g divides $f_0 - \varphi_0$, and since $\deg(f_0 - \varphi_0) < \deg g$ we see that $f_0 = \varphi_0$. Inductively, take the smallest integer i such that $f_i \neq \varphi_i$ (if such i exists). Dividing the above expression by g^i we find that g divides $f_i - \varphi_i$ and hence that such i cannot exist. This proves uniqueness.

We shall call the expression for f in terms of g in Theorem 5.3 the **g-adic expansion** of f. If $g(X) = X$, then the g-adic expansion is the usual expression of f as a polynomial.

Remark. In some sense, Theorem 5.2 redoes what was done in Theorem 8.1 of Chapter I for $\mathbf{Q}/\mathbf{Z}$; that is, express explicitly an element of K/A as a direct sum of its p-components.

§6. SYMMETRIC POLYNOMIALS

Let A be a commutative ring and let $t_1, \ldots, t_n$ be algebraically independent elements over A. Let X be a variable over $A[t_1, \ldots, t_n]$. We form the polynomial

$$\begin{aligned} F(X) &= (X - t_1) \cdots (X - t_n) \\ &= X^n - s_1 X^{n-1} + \cdots + (-1)^n s_n, \end{aligned}$$

where each $s_i = s_i(t_1, \ldots, t_n)$ is a polynomial in $t_1, \ldots, t_n$. Then for instance

$$s_1 = t_1 + \cdots + t_n \qquad \text{and} \qquad s_n = t_1 \cdots t_n.$$

The polynomials $s_1, \ldots, s_n$ are called the **elementary symmetric polynomials** of $t_1, \ldots, t_n$.

We leave it as an easy exercise to verify that s_i **is homogeneous of degree** i **in** $t_1, \ldots, t_n$.

Let σ be a permutation of the integers $(1, \ldots, n)$. Given a polynomial $f(t) \in A[t] = A[t_1, \ldots, t_n]$, we define σf to be

$$\sigma f(t_1, \ldots t_n) = f(t_{\sigma(1)}, \ldots, t_{\sigma(n)}).$$

If σ, τ are two permutations, then $\sigma\tau f = \sigma(\tau f)$ and hence the symmetric group G on n letters operates on the polynomial ring $A[t]$. A polynomial is called **symmetric** if $\sigma f = f$ for all $\sigma \in G$. It is clear that the set of symmetric polynomials is a subring of $A[t]$, which contains the constant polynomials

(i.e. A itself) and also contains the elementary symmetric polynomials $s_1, \dots, s_n$. We shall see below that $A[s_1, \dots, s_n]$ is the ring of symmetric polynomials.

Let $X_1, \dots, X_n$ be variables. We define the **weight** of a monomial

$$X_1^{v_1} \cdots X_n^{v_n}$$

to be $v_1 + 2v_2 + \cdots + nv_n$. We define the weight of a polynomial $g(X_1, \dots, X_n)$ to be the maximum of the weights of the monomials occurring in g.

Theorem 6.1. *Let $f(t) \in A[t_1, \dots, t_n]$ be symmetric of degree d. Then there exists a polynomial $g(X_1, \dots, X_n)$ of weight $\leqq d$ such that*

$$f(t) = g(s_1, \dots, s_n).$$

Proof. By induction on n. The theorem is obvious if $n = 1$, because $s_1 = t_1$.

Assume the theorem proved for polynomials in $n - 1$ variables.

If we substitute $t_n = 0$ in the expression for $F(X)$, we find

$$(X - t_1) \cdots (X - t_{n-1})X = X^n - (s_1)_0 X^{n-1} + \cdots + (-1)^{n-1}(s_{n-1})_0 X,$$

where $(s_i)_0$ is the expression obtained by substituting $t_n = 0$ in s_i. We see that $(s_1)_0, \dots, (s_{n-1})_0$ are precisely the elementary symmetric polynomials in $t_1, \dots, t_{n-1}$.

We now carry out induction on d. If $d = 0$, our assertion is trivial. Assume $d > 0$, and assume our assertion proved for polynomials of degree $< d$. Let $f(t_1, \dots, t_n)$ have degree d. There exists a polynomial $g_1(X_1, \dots, X_{n-1})$ of weight $\leqq d$ such that

$$f(t_1, \dots, t_{n-1}, 0) = g_1\big((s_1)_0, \dots, (s_{n-1})_0\big).$$

We note that $g_1(s_1, \dots, s_{n-1})$ has degree $\leqq d$ in $t_1, \dots, t_n$. The polynomial

$$f_1(t_1, \dots, t_n) = f(t_1, \dots, t_n) - g_1(s_1, \dots, s_{n-1})$$

has degree $\leqq d$ (in $t_1, \dots, t_n$) and is symmetric. We have

$$f_1(t_1, \dots, t_{n-1}, 0) = 0.$$

Hence f_1 is divisible by t_n, i.e. contains t_n as a factor. Since f_1 is symmetric, it contains $t_1 \cdots t_n$ as a factor. Hence

$$f_1 = s_n f_2(t_1, \dots, t_n)$$

for some polynomial f_2, which must be symmetric, and whose degree is

$\leqq d - n < d$. By induction, there exists a polynomial g_2 in n variables and weight $\leqq d - n$ such that

$$f_2(t_1, \dots, t_n) = g_2(s_1, \dots, s_n).$$

We obtain

$$f(t) = g_1(s_1, \dots, s_{n-1}) + s_n g_2(s_1, \dots, s_n),$$

and each term on the right has weight $\leqq d$. This proves our theorem.

We shall now prove that the elementary symmetric polynomials $s_1, \dots, s_n$ are algebraically independent over A.

If they are not, take a polynomial $f(X_1, \dots, X_n) \in A[X]$ of least degree and not equal to 0 such that

$$f(s_1, \dots, s_n) = 0.$$

Write f as a polynomial in X_n with coefficients in $A[X_1, \dots, X_{n-1}]$,

$$f(X_1, \dots, X_n) = f_0(X_1, \dots, X_{n-1}) + \cdots + f_d(X_1, \dots, X_{n-1})X_n^d.$$

Then $f_0 \neq 0$. Otherwise, we can write

$$f(X) = X_n \psi(X)$$

with some polynomial ψ, and hence $s_n \psi(s_1, \dots, s_n) = 0$. From this it follows that $\psi(s_1, \dots, s_n) = 0$, and ψ has degree smaller than the degree of f.

We substitute s_i for X_i in the above relation, and get

$$0 = f_0(s_1, \dots, s_{n-1}) + \cdots + f_d(s_1, \dots, s_{n-1})s_n^d.$$

This is a relation in $A[t_1, \dots, t_n]$, and we substitute 0 for t_n in this relation. Then all terms become 0 except the first one, which gives

$$0 = f_0\big((s_1)_0, \dots, (s_{n-1})_0\big),$$

using the same notation as in the proof of Theorem 6.1. This is a non-trivial relation between the elementary symmetric polynomials in $t_1, \dots, t_{n-1}$, a contradiction.

Example. (The Discriminant). Let $f(X) = (X - t_1) \cdots (X - t_n)$. Consider the product

$$\delta(t) = \prod_{i<j} (t_i - t_j).$$

For any permutation σ of $(1, \dots, n)$ we see at once that

$$\delta^\sigma(t) = \pm\delta(t).$$

Hence $\delta(t)^2$ is symmetric, and we call it the **discriminant**:

$$D_f = D(s_1, \ldots, s_n) = \prod_{i<j} (t_i - t_j)^2.$$

We thus view the discriminant as a polynomial in the elementary symmetric functions. For a continuation of the general theory, see §8. We shall now consider special cases.

Quadratic case. You should verify that for a quadratic polynomial $f(X) = X^2 + bX + c$, one has

$$D = b^2 - 4c.$$

Cubic case. Consider $f(X) = X^3 + aX + b$. We wish to prove that

$$D = -4a^3 - 27b^2.$$

Observe first that D is homogeneous of degree 6 in t_1, t_2. Furthermore, a is homogeneous of degree 2 and b is homogeneous of degree 3. By Theorem 6.1 we know that there exists some polynomial $g(X_2, X_3)$ of weight 6 such that $D = g(a, b)$. The only monomials $X_2^m X_3^n$ of weight 6, i.e. such that $2m + 3n = 6$ with integers m, $n \geqq 0$, are those for which $m = 3$, $n = 0$, or $m = 0$ and $n = 2$. Hence

$$g(X_2, X_3) = vX_2^3 + wX_3^2$$

where v, w are integers which must now be determined.

Observe that the integers v, w are universal, in the sense that for any special polynomial with special values of a, b its discriminant will be given by $g(a, b) = va^3 + wb^2$.

Consider the polynomial

$$f_1(X) = X(X-1)(X+1) = X^3 - X.$$

Then $a = -1$, $b = 0$, and $D = -va^3 = -v$. But also $D = 4$ by using the definition of the discriminant of the product of the differences of the roots, squared. Hence we get $v = -4$. Next consider the polynomial

$$f_2(X) = X^3 - 1.$$

Then $a = 0$, $b = -1$, and $D = 2b^2 = w$. But the three roots of f_2 are the cube roots of unity, namely

$$1, \frac{-1 + \sqrt{-3}}{2}, \frac{-1 - \sqrt{-3}}{2}.$$

Using the definition of the discriminant we find the value $D = -27$. Hence we get $w = -27$. This concludes the proof of the formula for the discriminant of the cubic when there is no X^2 term.

In general, consider a cubic polynomial

$$f(X) = X^3 - s_1X^2 + s_2X - s_3 = (X - t_1)(X - t_2)(X - t_3).$$

We find the value of the discriminant by reducing this case to the simpler case when there is no X^2 term. We make a translation, and let

$$Y = X - \tfrac{1}{3}s_1 \qquad \text{so} \quad X = Y + \tfrac{1}{3}s_1 = Y + \tfrac{1}{3}(t_1 + t_2 + t_3).$$

Then $f(X)$ becomes

$$f(X) = f^*(Y) = Y^3 + aY + b = (Y - u_1)(Y - u_2)(Y - u_3),$$

where $a = u_1u_2 + u_2u_3 + u_1u_3$ and $b = -u_1u_2u_3$, while $u_1 + u_2 + u_3 = 0$. We have

$$u_i = t_i - \tfrac{1}{3}s_1 \qquad \text{for} \quad i = 1, 2, 3,$$

and $u_i - u_j = t_i - t_j$ for all $i \neq j$, so the discriminant is unchanged, and you can easily get the formula in general. Do Exercise 12(b).

§7. MASON-STOTHERS THEOREM AND THE *abc* CONJECTURE

In the early 80s a new trend of thought about polynomials started with the discovery of an entirely new relation. Let $f(t)$ be a polynomial in one variable over the complex numbers if you wish (an algebraically closed field of characteristic 0 would do). We define

$$n_0(f) = \text{number of distinct roots of } f.$$

Thus $n_0(f)$ counts the zeros of f by giving each of them multiplicity 1, and $n_0(f)$ can be small even though $\deg f$ is large.

Theorem 7.1 (Mason-Stothers, [Mas 84], [Sto 81]). *Let $a(t)$, $b(t)$, $c(t)$ be relatively prime polynomials such that $a + b = c$. Then*

$$\max \deg\{a, b, c\} \leqq n_0(abc) - 1.$$

Proof. (Mason) Dividing by c, and letting $f = a/c$, $g = b/c$ we have

$$f + g = 1,$$

where f, g are rational functions. Differentiating we get $f' + g' = 0$, which we rewrite as

$$\frac{f'}{f}f + \frac{g'}{g}g = 0,$$

so that

$$\frac{b}{a} = \frac{g}{f} = -\frac{f'/f}{g'/g}.$$

Let

$$a(t) = c_1 \prod (t - \alpha_i)^{m_i}, \qquad b(t) = c_2 \prod (t - \beta_j)^{n_j}, \qquad c(t) = c_3 \prod (t - \gamma_k)^{r_k}.$$

Then by calculus algebraicized in Exercise 11(c), we get

$$\frac{b}{a} = -\frac{f'/f}{g'/g} = -\frac{\sum \frac{m_i}{t - \alpha_i} - \sum \frac{r_k}{t - \gamma_k}}{\sum \frac{n_j}{t - \beta_j} - \sum \frac{r_k}{t - \gamma_k}}.$$

A common denominator for f'/f and g'/g is given by the product

$$N_0 = \prod (t - \alpha_i) \prod (t - \beta_j) \prod (t - \gamma_k),$$

whose degree is $n_0(abc)$. Observe that $N_0 f'/f$ and $N_0 g'/g$ are both polynomials of degrees at most $n_0(abc) - 1$. From the relation

$$\frac{b}{a} = -\frac{N_0 f'/f}{N_0 g'/g},$$

and the fact that a, b are assumed relatively prime, we deduce the inequality in the theorem.

As an application, let us prove **Fermat's theorem** for polynomials. Thus let $x(t)$, $y(t)$, $z(t)$ be relatively prime polynomials such that one of them has degree $\geqq 1$, and such that

$$x(t)^n + y(t)^n = z(t)^n.$$

We want to prove that $n \leqq 2$. By the Mason-Stothers theorem, we get

$$n \deg x = \deg x(t)^n \leqq \deg x(t) + \deg y(t) + \deg z(t) - 1,$$

and similarly replacing x by y and z on the left-hand side. Adding, we find

$$n(\deg x + \deg y + \deg z) \leqq 3(\deg x + \deg y + \deg z) - 3.$$

This yields a contradiction if $n \geqq 3$.

As another application in the same vein, one has:

Davenport's theorem. *Let f, g be non-constant polynomials such that $f^3 - g^2 \neq 0$. Then*

$$\deg(f^3 - g^2) \geqq \tfrac{1}{2} \deg f - 1.$$

See Exercise 13.

One of the most fruitful analogies in mathematics is that between the integers **Z** and the ring of polynomials $F[t]$ over a field F. Evolving from the insights of Mason [Ma 84], Frey [Fr 87], Szpiro, and others, Masser and Oesterle formulated the *abc* conjecture for integers as follows. Let m be a non-zero integer. Define the **radical** of m to be

$$N_0(m) = \prod_{p|m} p,$$

i.e. the product of all the primes dividing m, taken with multiplicity 1.

The *abc* conjecture. *Given $\varepsilon > 0$, there exists a positive number $C(\varepsilon)$ having the following property. For any non-zero relative prime integers a, b, c such that $a + b = c$, we have*

$$\max(|a|, |b|, |c|) \leqq C(\varepsilon)N_0(abc)^{1+\varepsilon}.$$

Observe that the inequality says that many prime factors of a, b, c occur to the first power, and that if "small" primes occur to high powers, then they have to be compensated by "large" primes occurring to the first power. For instance, one might consider the equation

$$2^n \pm 1 = m.$$

For m large, the *abc* conjecture would state that m has to be divisible by large primes to the first power. This phenomenon can be seen in the tables of [BLSTW 83].

Stewart–Tijdeman [ST 86] have shown that it is necessary to have the ε in the formulation of the conjecture. Subsequent examples were communicated to me by Wojtek Jastrzebowski and Dan Spielman as follows.

We have to give examples such that for all $C > 0$ there exist natural numbers a, b, c relatively prime such that $a + b = c$ and $|a| > CN_0(abc)$. But trivially,

$$2^n | (3^{2^n} - 1).$$

We consider the relations $a_n + b_n = c_n$ given by

$$3^{2^n} - 1 = c_n.$$

It is clear that these relations provide the desired examples. Other examples can be constructed similarly, since the role of 3 and 2 can be played by other integers. Replace 2 by some prime, and 3 by an integer $\equiv 1 \bmod p$.

The *abc* conjecture implies what we shall call the

Asymptotic Fermat Theorem. *For all n sufficiently large, the equation*

$$x^n + y^n = z^n$$

has no solution in relatively prime integers $\neq 0$.

The proof follows exactly the same pattern as for polynomials, except that we write things down multiplicatively, and there is a $1 + \varepsilon$ floating around. The extent to which the *abc* conjecture will be proved with an explicit constant $C(\varepsilon)$ (or say $C(1)$ to fix ideas) yields the corresponding explicit determination of the bound for n in the application. We now go into other applications.

Hall's conjecture [Ha 71]. *If u, v are relatively prime non-zero integers such that $u^3 - v^2 \neq 0$, then*

$$|u^3 - v^2| \gg |u|^{1/2-\varepsilon}.$$

The symbol $\gg$ means that the left-hand side is $\geqq$ the right-hand side times a constant depending only on ε. Again the proof is immediate from the *abc* conjecture. Actually, the hypothesis that u, v are relatively prime is not necessary; the general case can be reduced to the relatively prime case by extracting common factors, and Hall stated his conjecture in this more general way. However, he also stated it without the epsilon in the exponent, and that does not work, as was realized later. As in the polynomial case, Hall's conjecture describes how small $|u^3 - v^2|$ can be, and the answer is not too small, as described by the right-hand side.

The Hall conjecture can also be interpreted as giving a bound for integral relatively prime solutions of

$$v^2 = u^3 + b \quad \text{with integral } b.$$

Then we find

$$|u| \ll |b|^{2+\varepsilon}.$$

More generally, in line with conjectured inequalities from Lang–Waldschmidt [La 78], let us fix non-zero integers A, B and let u, v, k, m, n be variable, with u, v relatively prime and $mv > m + n$. Put

$$Au^m + Bv^n = k.$$

By the *abc* conjecture, one derives easily that

$$(1) \qquad |u| \ll N_0(k)^{\frac{m}{mn-(m+n)}(1+\varepsilon)} \quad \text{and} \quad |v| \ll N_0(k)^{\frac{mn}{mn-(m+n)}(1+\varepsilon)}.$$

From this one gets

$$|k| \ll N_0(k)^{\frac{mn}{mn-(m+n)}(1+\varepsilon)}.$$

The Hall conjecture is a special case after we replace $N_0(k)$ with $|k|$, because $N_0(k) \leqq |k|$.

Next take $m = 3$ and $n = 2$, but take $A = 4$ and $B = -27$. In this case we write

$$D = 4u^3 - 27v^2$$

and we get

(2) $$|u| \ll N_0(D)^{2+\varepsilon} \quad \text{and} \quad |v| \ll N_0(D)^{3+\varepsilon}.$$

These inequalities are supposed to hold at first for u, v relatively prime. Suppose we allow u, v to have some bounded common factor, say d. Write

$$u = u'd \quad \text{and} \quad v = v'd$$

with u', v' relatively prime. Then

$$D = 4d^3u'^3 - 27d^2v'^2.$$

Now we can apply inequality (1) with $A = 4d^3$ and $B = -27d^2$, and we find the same inequalities (2), with the constant implicit in the sign $\ll$ depending also on d, or on some fixed bound for such a common factor. Under these circumstances, we call inequalities (2) the **generalized Szpiro conjecture**.

The original Szpiro conjecture was stated in a more sophisticated situation, cf. [La 90] for an exposition, and Szpiro's inequality was stated in the form

$$|D| \ll N(D)^{6+\varepsilon},$$

where $N(D)$ is a more subtle invariant, but for our purposes, it is sufficient and much easier to use the radical $N_0(D)$.

The point of D is that it occurs as a discriminant. The trend of thoughts in the direction we are discussing was started by Frey [Fr 87], who associated with each solution of $a + b = c$ the polynomial

$$x(x - a)(x + b),$$

which we call the **Frey polynomial**. (Actually Frey associated the curve defined by the equation $y^2 = x(x - a)(x + b)$, for much deeper reasons, but only the polynomial on the right-hand side will be needed here.) The discriminant of the polynomial is the product of the differences of the roots squared, and so

$$D = (abc)^2.$$

We make a translation

$$\xi = x + \frac{b - a}{3}$$

to get rid of the x^2-term, so that our polynomial can be rewritten

$$\xi^3 - \gamma_2\xi - \gamma_3,$$

where γ_2, γ_3 are homogeneous in a, b of appropriate weight. The discriminant does not change because the roots of the polynomial in ξ are

translations of the roots of the polynomial in x. Then

$$D = 4\gamma_2^3 - 27\gamma_3^2.$$

The translation with $(b - a)/3$ introduces a small denominator. One may avoid this denominator by using the polynomial $x(x - 3a)(x - 3b)$, so that γ_2, γ_3 then come out to be integers, and one can apply the generalized Szpiro conjecture to the discriminant, which then has an extra factor $D = 3^6(abc)^2$.

It is immediately seen that the generalized Szpiro conjecture implies asymptotic Fermat. Conversely:

Generalized Szpiro implies the abc conjecture.

Indeed, the correspondence $(a, b) \leftrightarrow (\gamma_2, \gamma_3)$ is invertible, and has the "right" weight. A simple algebraic manipulation shows that the generalized Szpiro estimates on γ_2, γ_3 imply the desired estimates on $|a|, |b|$. (Do Exercise 14.) From the equivalence between *abc* and generalized Szpiro, one can use the examples given earlier to show that the epsilon is needed in the Szpiro conjecture.

Finally, note that the polynomial case of the Mason-Stothers theorem and the case of integers are not independent, or specifically the Davenport theorem and Hall's conjecture are related. Examples in the polynomial case parametrize cases with integers when we substitute integers for the variables. Such examples are given in [BCHS 65], one of them (due to Birch) being

$$f(t) = t^6 + 4t^4 + 10t^2 + 6 \quad \text{and} \quad g(t) = t^9 + 6t^7 + 21t^5 + 35t^3 + \tfrac{63}{2}t,$$

whence

$$\deg(f(t)^3 - g(t)^2) = \tfrac{1}{2}\deg f + 1.$$

This example shows that Davenport's inequality is best possible, because the degree attains the lowest possible value permissible under the theorem. Substituting large integral values of $t \equiv 2 \bmod 4$ gives examples of similarly low values for $x^3 - y^2$. For other connections of all these matters, cf. [La 90].

Bibliography

[BCHS 65] B. Birch, S. Chowla, M. Hall, and A. Schinzel, On the difference $x^3 - y^2$, *Norske Vid. Selsk. Forrh.* **38** (1965) pp. 65–69

[BLSTW 83] J. Brillhart, D. H. Lehmer, J. L. Selfridge, B. Tuckerman, and S. Wagstaff Jr., Factorization of $b^n \pm 1$, $b = 2, 3, 5, 6, 7, 10, 11$ up to high powers, *Contemporary Mathematics* Vol. **22**, AMS, Providence, RI, 1983

[Dav 65] H. Davenport, On $f^3(t) - g^2(t)$, *Norske Vid. Selsk. Forrh.* **38** (1965) pp. 86–87

[Fr 87] G. Frey, Links between solutions of $A - B = C$ and elliptic curves, *Number Theory, Lecture Notes* **1380**, Springer-Verlag, New York, 1989 pp. 31–62

[Ha 71] M. HALL, The diophantine equation $x^3 - y^2 = k$, *Computers and Number Theory*, ed. by A. O. L. Atkin and B. Birch, Academic Press, London 1971 pp. 173–198

[La 90] S. LANG, Old and new conjectured diophantine inequalities, *Bull. AMS* Vol. **23** No. 1 (1990) pp. 37–75

[Ma 84a] R. C. MASON, Equations over function fields, Springer Lecture Notes **1068** (1984), pp. 149–157; in *Number Theory, Proceedings of the Noordwijkerhout*, 1983

[Ma 84b] R. C. MASON, Diophantine equations over function fields, *London Math. Soc. Lecture Note Series* Vol. **96**, Cambridge University Press, Cambridge, 1984

[Ma 84c] R. C. MASON, The hyperelliptic equation over function fields, *Math. Proc. Cambridge Philos. Soc.* **93** (1983) pp. 219–230

[Si 88] J. SILVERMAN, Wieferich's criterion and the *abc* conjecture, *Journal of Number Theory* **30** (1988) pp. 226–237

[ST 86] C. L. STEWART and R. TIJDEMAN, On the Oesterle–Masser Conjecture, *Mon. Math.* **102** (1986) pp. 251–257

See additional references at the end of the chapter.

§8. THE RESULTANT

In this section, we assume that the reader is familiar with determinants. The theory of determinants will be covered later. The section can be viewed as giving further examples of symmetric functions.

Let A be a commutative ring and let $v_0, \dots, v_n$, $w_0, \dots, w_m$ be algebraically independent over A. We form two polynomials:

$$f_v(X) = v_0 X^n + \cdots + v_n,$$

$$g_w(X) = w_0 X^m + \cdots + w_m.$$

We define the **resultant** of (v, w), or of f_v, g_w, to be the determinant

$$\begin{array}{l} m\left\{\begin{array}{l}\vphantom{v_0}\\ \vphantom{v_0}\\ \vphantom{v_0}\\ \vphantom{v_0}\end{array}\right. \\ n\left\{\begin{array}{l}\vphantom{w_0}\\ \vphantom{w_0}\\ \vphantom{w_0}\\ \vphantom{w_0}\end{array}\right. \end{array} \underbrace{\begin{vmatrix} v_0 v_1 \cdots v_n & & \\ & v_0 v_1 \cdots v_n & \\ & \cdots\cdots\cdots\cdots & \\ & & v_0 v_1 \cdots v_n \\ w_0 w_1 \cdots w_m & & \\ & w_0 w_1 \cdots w_m & \\ & \cdots\cdots\cdots\cdots & \\ & & w_0 w_1 \cdots w_m \end{vmatrix}}_{m+n} .$$

The blank spaces are supposed to be filled with zeros.

If we substitute elements $(a) = (a_0, \ldots, a_n)$ and $(b) = (b_0, \ldots, b_m)$ in A for (v) and (w) respectively in the coefficients of f_v and g_w, then we obtain polynomials f_a and g_b with coefficients in A, and *we define their* **resultant** *to be the determinant obtained by substituting* (a) *for* (v) *and* (b) *for* (w) *in the determinant.* We shall write the resultant of f_v, g_w in the form

$$\operatorname{Res}(f_v, g_w) \quad \text{or} \quad R(v, w).$$

The resultant $\operatorname{Res}(f_a, g_b)$ is then obtained by substitution of (a), (b) for (v), (w) respectively.

We observe that $R(v, w)$ is a polynomial with integer coefficients, i.e. we may take $A = \mathbf{Z}$. If z is a variable, then

$$R(zv, w) = z^m R(v, w) \qquad \text{and} \qquad R(v, zw) = z^n R(v, w)$$

as one sees immediately by factoring out z from the first m rows (resp. the last n rows) in the determinant. Thus R is homogeneous of degree m in its first set of variables, and homogeneous of degree n in its second set of variables. Furthermore, $R(v, w)$ contains the monomial

$$v_0^m w_m^n$$

with coefficient 1, when expressed as a sum of monomials.

If we substitute 0 for v_0 and w_0 in the resultant, we obtain 0, because the first column of the determinant vanishes.

Let us work over the integers $\mathbf{Z}$. We consider the linear equations

$$\begin{aligned}
X^{m-1} f_v(X) &= v_0 X^{n+m-1} + v_1 X^{n+m-2} + \cdots + v_n X^{m-1} \\
X^{m-2} f_v(X) &= \qquad\qquad\quad v_0 X^{n+m-2} + \cdots \quad + v_n X^{m-2} \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
f_v(X) &= \qquad\qquad\qquad\qquad v_0 X^n + \cdots \qquad + v_n \\
X^{n-1} g_w(X) &= w_0 X^{n+m-1} + w_1 X^{n+m-2} + \cdots + w_m X^{n-1} \\
X^{n-2} g_w(X) &= \qquad\qquad\quad w_0 X^{n+m-2} + \cdots \quad + w_m X^{n-2} \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
g_w(X) &= \qquad\qquad\qquad\qquad w_0 X^m + \cdots \qquad + w_m.
\end{aligned}$$

Let C be the column vector on the left-hand side, and let

$$C_0, \ldots, C_{m+n}$$

be the column vectors of coefficients. Our equations can be written

$$C = X^{n+m-1} C_0 + \cdots + 1 \cdot C_{m+n}.$$

By Cramer's rule, applied to the last coefficient which is $= 1$,

$$R(v, w) = \det(C_0, \ldots, C_{m+n}) = \det(C_0, \ldots, C_{m+n-1}, C).$$

From this we see that there exist polynomials $\varphi_{v,w}$ and $\psi_{v,w}$ in $\mathbf{Z}[v, w][X]$ such that

$$\varphi_{,v,w} f_v + \psi_{v,w} g_w = R(v,w) = \operatorname{Res}(f_v, f_w).$$

Note that $R(v, w) \in \mathbf{Z}[v, w]$ but that the polynomials on the left-hand side involve the variable X.

If $\lambda: \mathbf{Z}[v, w] \to A$ is a homomorphism into a commutative ring A and we let $\lambda(v) = (a)$, $\lambda(w) = (b)$, then

$$\varphi_{a,b} f_a + \psi_{a,b} g_b = R(a, b) = \operatorname{Res}(f_a, f_b).$$

Thus from the universal relation of the resultant over $\mathbf{Z}$ we obtain a similar relation for every pair of polynomials, in any commutative ring A.

Proposition 8.1. *Let K be a subfield of a field L, and let f_a, g_b be polynomials in $K[X]$ having a common root ξ in L. Then $R(a, b) = 0$.*

Proof. If $f_a(\xi) = g_b(\xi) = 0$, then we substitute ξ for X in the expression obtained for $R(a, b)$ and find $R(a, b) = 0$.

Next, we shall investigate the relationship between the resultant and the roots of our polynomials f_v, g_w. We need a lemma.

Lemma 8.2. *Let $h(X_1, \dots, X_n)$ be a polynomial in n variables over the integers $\mathbf{Z}$. If h has the value 0 when we substitute X_1 for X_2 and leave the other X_i fixed ($i \neq 2$), then $h(X_1, \dots, X_n)$ is divisible by $X_1 - X_2$ in $\mathbf{Z}[X_1, \dots, X_n]$.*

Proof. Exercise for the reader.

Let $v_0, t_1, \dots, t_n, w_0, u_1, \dots, u_m$ be algebraically independent over $\mathbf{Z}$ and form the polynomials

$$f_v = v_0(X - t_1) \cdots (X - t_n) = v_0 X^n + \cdots + v_n,$$

$$g_w = w_0(X - u_1) \cdots (X - u_m) = w_0 X^m + \cdots + w_m.$$

Thus we let

$$v_i = (-1)^i v_0 s_i(t) \qquad \text{and} \qquad w_j = (-1)^j w_0 s_j(u).$$

We leave to the reader the easy verification that

$$v_0, v_1, \dots, v_n, w_0, w_1, \dots, w_m$$

are algebraically independent over $\mathbf{Z}$.

Proposition 8.3. *Notation being as above, we have*

$$\operatorname{Res}(f_v, g_w) = v_0^m w_0^n \prod_{i=1}^{n} \prod_{j=1}^{m} (t_i - u_j).$$

Proof. Let S be the expression on the right-hand side of the equality in the statement of the proposition.

Since $R(v, w)$ is homogeneous of degree m in its first variables, and homogeneous of degree n in its second variables, it follows that

$$R = v_0^m w_0^n h(t, u)$$

where $h(t, u) \in \mathbf{Z}[t, u]$. By Proposition 8.1, the resultant vanishes when we substitute t_i for u_j ($i = 1, \ldots, n$ and $j = 1, \ldots, m$), whence by the lemma, viewing R as an element of $\mathbf{Z}[v_0, w_0, t, u]$ it follows that R is divisible by $t_i - u_j$ for each pair (i, j). Hence S divides R in $\mathbf{Z}[v_0, w_0, t, u]$, because $t_i - u_j$ is obviously a prime in that ring, and different pairs (i, j) give rise to different primes.

From the product expression for S, namely

$$S = v_0^m w_0^n \prod_{i=1}^{n} \prod_{j=1}^{m} (t_i - u_j), \tag{1}$$

we obtain

$$\prod_{i=1}^{n} g(t_i) = w_0^n \prod_{i=1}^{n} \prod_{j=1}^{m} (t_i - u_j),$$

whence

$$S = v_0^m \prod_{i=1}^{n} g(t_i). \tag{2}$$

Similarly,

$$S = (-1)^{nm} w_0^n \prod_{j=1}^{m} f(u_j). \tag{3}$$

From (2) we see that S is homogeneous and of degree n in (w), and from (3) we see that S is homogeneous and of degree m in (v). Since R has exactly the same homogeneity properties, and is divisible by S, it follows that $R = cS$ for some integer c. Since both R and S have a monomial $v_0^m w_m^n$ occurring in them with coefficient 1, it follows that $c = 1$, and our proposition is proved.

We also note that the three expressions found for S above now give us a factorization of R. We also get a converse for Proposition 8.1.

Corollary 8.4. *Let f_a, g_b be polynomials with coefficients in a field K, such that $a_0 b_0 \neq 0$, and such that f_a, g_b split in factors of degree 1 in $K[X]$. Then $\operatorname{Res}(f_a, g_b) = 0$ if and only if f_a and g_b have a root in common.*

Proof. Assume that the resultant is 0. If

$$f_a = a_0(X - \alpha_1) \cdots (X - \alpha_n),$$

$$g_b = b_0(X - \beta_1) \cdots (X - \beta_n),$$

is the factorization of f_a, g_b, then we have a homomorphism

$$\mathbf{Z}[v_0, t, w_0, u] \to K$$

such that $v_0 \mapsto a_0$, $w_0 \mapsto b_0$, $t_i \mapsto \alpha_i$, and $u_j \mapsto \beta_j$ for all i, j. Then

$$0 = \operatorname{Res}(f_a, g_b) = a_0^m b_0^n \prod_i \prod_j (\alpha_i - \beta_j),$$

whence f_a, f_b have a root in common. The converse has already been proved.

We deduce one more relation for the resultant in a special case. Let f_v be as above,

$$f_v(X) = v_0 X^n + \cdots + v_n = v_0(X - t_1) \cdots (X - t_n).$$

From (2) we know that if f_v' is the derivative of f_v, then

$$\operatorname{Res}(f_v, f_v') = v_0^{n-1} \prod_i f'(t_i). \tag{4}$$

Using the product rule for differentiation, we find:

$$f_v'(X) = \sum_i v_0(X - t_1) \cdots \widehat{(X - t_i)} \cdots (X - t_n),$$

$$f_v'(t_i) = v_0(t_i - t_1) \cdots \widehat{(t_i - t_i)} \cdots (t_i - t_n),$$

where a roof over a term means that this term is to be omitted.

We define the **discriminant** of f_v to be

$$D(f_v) = D(v) = (-1)^{n(n-1)/2} v_0^{2n-2} \prod_{i \neq j} (t_i - t_j).$$

Proposition 8.5. *Let f_v be as above and have algebraically independent coefficients over* $\mathbf{Z}$. *Then*

$$\operatorname{Res}(f_v, f_v') = v_0^{2n-1} \prod_{i \neq j} (t_i - t_j) = (-1)^{n(n-1)/2} v_0 D(f_v). \tag{5}$$

Proof. One substitutes the expression obtained for $f_v'(t_i)$ into the product (4). The result follows at once.

When we substitute 1 for v_0, we find that the discriminant as we defined it in the preceding section coincides with the present definition. In particular, we find an explicit formula for the discriminant. The formulas in the special case of polynomials of degree 2 and 3 will be given as exercises.

Note that the discriminant can also be written as the product

$$D(f_v) = v_0^{2n-2} \prod_{i<j} (t_i - t_j)^2.$$

Serre once pointed out to me that the sign $(-1)^{n(n-1)/2}$ was missing in the first edition of this book, and that this sign error is quite common in the literature, occurring as it does in van der Waerden, Samuel, and Hilbert (but not in his collected works, corrected by Olga Taussky); on the other hand the sign is correctly given in Weber's *Algebra*, Vol. I, 50.

For a continuation of this section, see Chapter IX, §3 and §4.

§9. POWER SERIES

Let X be a letter, and let G be the monoid of functions from the set $\{X\}$ to the natural numbers. If $\nu \in \mathbf{N}$, we denote by X^ν the function whose value at X is ν. Then G is a multiplicative monoid, already encountered when we discussed polynomials. Its elements are $X^0, X^1, X^2, \ldots, X^\nu, \ldots$.

Let A be a commutative ring, and let $A[[X]]$ be the set of functions from G into A, without any restriction. Then an element of $A[[X]]$ may be viewed as assigning to each monomial X^ν a coefficient $a_\nu \in A$. We denote this element by

$$\sum_{\nu=0}^{\infty} a_\nu X^\nu.$$

The summation symbol is not a sum, of course, but we shall write the above expression also in the form

$$a_0 X^0 + a_1 X^1 + \cdots$$

and we call it a **formal power series** with coefficients in A, in one variable. We call $a_0, a_1, \ldots$ its coefficients.

Given two elements of $A[[X]]$, say

$$\sum_{\nu=0}^{\infty} a_\nu X^\nu \quad \text{and} \quad \sum_{\mu=0}^{\infty} b_\mu X^\mu,$$

we define their product to be

$$\sum_{i=0}^{\infty} c_i X^i$$

where

$$c_i = \sum_{\nu+\mu=i} a_\nu b_\mu.$$

Just as with polynomials, one defines their sum to be

$$\sum_{\nu=0}^{\infty} (a_\nu + b_\nu) X^\nu.$$

Then we see that the power series form a ring, the proof being the same as for polynomials.

One can also construct the power series ring in several variables $A[[X_1, \ldots, X_n]]$ in which every element can be expressed in the form

$$\sum_{(\nu)} a_{(\nu)} X_1^{\nu_1} \cdots X_n^{\nu_n} = \sum a_{(\nu)} M_{(\nu)}(X_1, \ldots, X_n)$$

with unrestricted coefficients $a_{(\nu)}$ in bijection with the n-tuples of integers $(\nu_1, \ldots, \nu_n)$ such that $\nu_i \geqq 0$ for all i. It is then easy to show that there is an isomorphism between $A[[X_1, \ldots, X_n]]$ and the repeated power series ring $A[[X_1]] \cdots [[X_n]]$. We leave this as an exercise for the reader.

The next theorem will give an analogue of the Euclidean algorithm for power series. However, instead of dealing with power series over a field, it is important to have somewhat more general coefficients for certain applications, so we have to introduce a little more terminology.

Let A be a ring and I an ideal. We assume that

$$\bigcap_{\nu=1}^{\infty} I^{\nu} = \{0\}.$$

We can view the powers I^{ν} as defining neighborhoods of 0 in A, and we can transpose the usual definition of Cauchy sequence in analysis to this situation, namely: we define a sequence $\{a_n\}$ in A to be **Cauchy** if given some power I^{ν} there exists an integer N such that for all $m, n \geqq N$ we have

$$a_m - a_n \in I^{\nu}.$$

Thus I^{ν} corresponds to the given ϵ of analysis. Then we have the usual notion of **convergence** of a sequence to an element of A. One says that A is **complete in the I-adic topology** if every Cauchy sequence converges.

Perhaps the most important example of this situation is when A is a local ring and $I = \mathfrak{m}$ is its maximal ideal. By a **complete local ring**, one always means a local ring which is complete in the $\mathfrak{m}$-adic topology.

Let k be a field. Then the power series ring

$$R = k[[X_1, \ldots, X_n]]$$

in n variables is such a complete local ring. Indeed, let $\mathfrak{m}$ be the ideal generated by the variables $X_1, \ldots, X_n$. Then $R/\mathfrak{m}$ is naturally isomorphic to the field k itself, so $\mathfrak{m}$ is a maximal ideal. Furthermore, any power series of the form

$$f(X) = c_0 - f_1(X)$$

with $c_0 \in k$, $c_0 \neq 0$ and $f_1(X) \in \mathfrak{m}$ is invertible. To prove this, one may first assume without loss of generality that $c_0 = 1$. Then

$$(1 - f_1(X))^{-1} = 1 + f_1(X) + f_1(X)^2 + f_1(X)^3 + \cdots$$

gives the inverse. Thus we see that $\mathfrak{m}$ is the unique maximal ideal and R is local. It is immediately verified that R is complete in the sense we have just defined. The same argument shows that if k is not a field but c_0 is invertible in k, then again $f(X)$ is invertible.

Again let A be a ring. We may view the power series ring in n variables ($n > 1$) as the ring of power series in one variable X_n over the ring of power series in $n - 1$ variables, that is we have a natural identification

$$A[[X_1, \ldots, X_n]] = A[[X_1, \ldots, X_{n-1}]][[X_n]].$$

If $A = k$ is a field, the ring $k[[X_1, \ldots, X_{n-1}]]$ is then a complete local ring. More generally, if $\mathfrak{o}$ is a complete local ring, then the power series ring $\mathfrak{o}[[X]]$ is a complete local ring, whose maximal ideal is $(\mathfrak{m}, X)$ where $\mathfrak{m}$ is the maximal ideal of $\mathfrak{o}$. Indeed, if a power series $\sum a_{\nu} X^{\nu}$ has unit constant

term $a_0 \in \mathfrak{o}^*$, then the power series is a unit in $\mathfrak{o}[[X]]$, because first, without loss of generality, we may assume that $a_0 = 1$, and then we may invert $1 + h$ with $h \in (\mathfrak{m}, X)$ by the geometric series $1 - h + h^2 - h^3 + \cdots$.

In a number of problems, it is useful to reduce certain questions about power series in several variables over a field to questions about power series in one variable over the more complicated ring as above. We shall now apply this decomposition to the Euclidean algorithm for power series.

Theorem 9.1. *Let $\mathfrak{o}$ be a complete local ring with maximal ideal $\mathfrak{m}$. Let*

$$f(X) = \sum_{i=0}^{\infty} a_i X^i$$

be a power series in $\mathfrak{o}[[X]]$ (one variable), such that not all a_i lie in $\mathfrak{m}$. Say $a_0, \ldots, a_{n-1} \in \mathfrak{m}$, and $a_n \in \mathfrak{o}^*$ is a unit. Given $g \in \mathfrak{o}[[X]]$ we can solve the equation

$$g = qf + r$$

uniquely with $q \in \mathfrak{o}[[X]]$, $r \in \mathfrak{o}[X]$, and $\deg r \leqq n - 1$.

Proof (Manin). Let α and τ be the projections on the beginning and tail end of the series, given by

$$\alpha\colon \sum b_i X^i \mapsto \sum_{i=0}^{n-1} b_i X^i = b_0 + b_1 X + \cdots + b_{n-1} X^{n-1},$$

$$\tau\colon \sum b_i X^i \mapsto \sum_{i=n}^{\infty} b_i X^{i-n} = b_n + b_{n+1} X + b_{n+2} X^2 + \cdots.$$

Note that $\tau(hX^n) = h$ for any $h \in \mathfrak{o}[[X]]$; and h is a polynomial of degree $< n$ if and only if $\tau(h) = 0$.

The existence of q, r is equivalent with the condition that there exists q such that

$$\tau(g) = \tau(qf).$$

Hence our problem is equivalent with solving

$$\tau(g) = \tau(q\alpha(f)) + \tau(q\tau(f)X^n) = \tau(q\alpha(f)) + q\tau(f).$$

Note that $\tau(f)$ is invertible. Put $Z = q\tau(f)$. Then the above equation is equivalent with

$$\tau(g) = \tau\left(Z\frac{\alpha(f)}{\tau(f)}\right) + Z = \left(I + \tau \circ \frac{\alpha(f)}{\tau(f)}\right)Z.$$

Note that

$$\tau \circ \frac{\alpha(f)}{\tau(f)}\colon \mathfrak{o}[[X]] \to \mathfrak{m}\mathfrak{o}[[X]],$$

because $\alpha(f)/\tau(f) \in \mathfrak{m}\mathfrak{o}[[X]]$. We can therefore invert to find Z, namely

$$Z = \left(I + \tau \circ \frac{\alpha(f)}{\tau(f)}\right)^{-1} \tau(g),$$

which proves both existence and uniqueness and concludes the proof.

Theorem 9.2. (Weierstrass Preparation). *The power series f in the previous theorem can be written uniquely in the form*

$$f(X) = (X^n + b_{n-1}X^{n-1} + \cdots + b_0)u,$$

where $b_i \in \mathfrak{m}$, and u is a unit in $\mathfrak{o}[[X]]$.

Proof. Write uniquely

$$X^n = qf + r,$$

by the Euclidean algorithm. Then q is invertible, because

$$q = c_0 + c_1 X + \cdots,$$

$$f = \cdots + a_n X^n + \cdots,$$

so that

$$1 \equiv c_0 a_n \pmod{\mathfrak{m}},$$

and therefore c_0 is a unit in $\mathfrak{o}$. We obtain $qf = X^n - r$, and

$$f = q^{-1}(X^n - r),$$

with $r \equiv 0 \pmod{\mathfrak{m}}$. This proves the existence. Uniqueness is immediate.

The integer n in Theorems 9.1 and 9.2 is called the **Weierstrass degree** of f, and is denoted by $\deg_W f$. We see that a power series not all of whose coefficients lie in $\mathfrak{m}$ can be expressed as a product of a polynomial having the given Weierstrass degree, times a unit in the power series ring. Furthermore, all the coefficients of the polynomial except the leading one lie in the maximal ideal. Such a polynomial is called **distinguished**, or a **Weierstrass polynomial.**

Remark. I rather like the use of the Euclidean algorithm in the proof of the Weierstrass Preparation theorem. However, one can also give a direct proof exhibiting explicitly the recursion relations which solve for the coefficients of u, as follows. Write $u = \sum c_i X^i$. Then we have to solve the equations

$$b_0 c_0 = a_0,$$

$$b_0 c_1 + b_1 c_0 = a_1,$$

$$\cdots$$

$$b_0 c_{n-1} + \cdots + b_{n-1} c_0 = a_{n-1},$$

$$b_0 c_n + \cdots + c_0 = a_n,$$

$$b_0 c_{n+1} + \cdots + c_1 = a_{n+1},$$

$$\cdots$$

In fact, the system of equations has a unique solution mod $\mathfrak{m}^r$ for each positive integer r, after selecting c_0 to be a unit, say $c_0 = 1$. Indeed, from the first n equations (from 0 to $n-1$) we see that $b_0, \ldots, b_{n-1}$ are uniquely determined to be 0 mod $\mathfrak{m}$. Then $c_n, c_{n+1}, \ldots$ are uniquely determined mod $\mathfrak{m}$ by the subsequent equations. Now inductively, suppose we have shown that the coefficients b_i, c_j are uniquely determined mod $\mathfrak{m}^r$. Then one sees immediately that from the conditions $a_0, \ldots, a_{n-1} \equiv 0$ mod $\mathfrak{m}$ the first n equations define b_i uniquely mod $\mathfrak{m}^{r+1}$ because all $b_i \equiv 0$ mod $\mathfrak{m}$. Then the subsequent equations define c_j mod $\mathfrak{m}^{r+1}$ uniquely from the values of b_i mod $\mathfrak{m}^{r+1}$ and c_j mod $\mathfrak{m}^r$. The unique system of solutions mod $\mathfrak{m}^r$ for each r then defines a solution in the projective limit, which is the complete local ring.

We now have all the tools to deal with unique factorization in one important case.

Theorem 9.3. *Let k be a field. Then $k[[X_1, \ldots, X_n]]$ is factorial.*

Proof. Let $f(x) = f(X_1, \ldots, X_n) \in k[[X]]$ be $\neq 0$. After making a sufficiently general linear change of variables (when k is infinite)

$$x_i = \sum c_{ij} Y_j \quad \text{with} \quad c_{ij} \in k,$$

we may assume without loss of generality that $f(0, \ldots, 0, x_n) \neq 0$. (When k is finite, one has to make a non-linear change, cf. Theorem 2.1 of Chapter VIII.) Indeed, if we write $f(X) = f_d(X) +$ higher terms, where $f_d(X)$ is a homogeneous polynomial of degree $d \geqq 0$, then changing the variables as above preserves the degree of each homogeneous component of f, and since k is assumed infinite, the coefficients c_{ij} can be taken so that in fact each power Y_i^d ($i = 1, \ldots, n$) occurs with non-zero coefficient.

We now proceed by induction on n. Let $R_n = k[[X_1, \ldots, X_n]]$ be the power series in n variables, and assume by induction that R_{n-1} is factorial. By Theorem 9.2, write $f = gu$ where u is a unit and g is a Weierstrass polynomial in $R_{n-1}[X_n]$. By Theorem 2.3, $R_{n-1}[X_n]$ is factorial, and so we can write g as a product of irreducible elements $g_1, \ldots, g_r \in R_{n-1}[X_n]$, so $f = g_1 \cdots g_r u$, where the factors g_i are uniquely determined up to multiplication by units. This proves the existence of a factorization. As to uniqueness, suppose f is expressed as a product of irreducible elements in R_n, $f = f_1 \cdots f_s$. Then $f_q(0, \ldots, 0, x_n) \neq 0$ for each $q = 1, \ldots, s$, so we can write $f_q = h_q u_q'$ where u_q' is a unit and h_q is a Weierstrass polynomial, necessarily irreducible in $R_{n-1}[X_n]$. Then $f = gu = \prod h_q \prod u_q'$ with g and all h_q Weierstrass polynomials. By Theorem 9.2, we must have $g = \prod h_q$, and since $R_{n-1}[X_n]$ is factorial, it follows that the polynomials h_q are the same as the polynomials g_i, up to units. This proves uniqueness.

Remark. As was pointed out to me by Dan Anderson, I incorrectly stated in a previous printing that if $\mathfrak{O}$ is a factorial complete local ring, then $\mathfrak{O}[[X]]$ is also factorial. This assertion is false, as shown by the example

$$k(t)[[X_1, X_2, X_3]]/(X_1^2 + X_2^2 + X_3^2)$$

due to P. Salmon, *Su un problema post da P. Samuel*, Atti Acad. Naz. Lincei Rend. Cl. Sc. Fis. Matem. **40(8)** (1966) pp. 801–803. It is true that if $\mathfrak{O}$ is a regular local ring *in addition* to being complete, then $\mathfrak{O}[[X]]$ is factorial, but this is a deeper theorem. The simple proof I gave for the power series over a field is classical. I chose the exposition in [GrH 78].

Theorem 9.4. *If A is Noetherian, then $A[[X]]$ is also Noetherian.*

Proof. Our argument will be a modification of the argument used in the proof of Hilbert's theorem for polynomials. We shall consider elements of lowest degree instead of elements of highest degree.

Let $\mathfrak{A}$ be an ideal of $A[[X]]$. We let $\mathfrak{a}_i$ be the set of elements $a \in A$ such that a is the coefficient of X^i in a power series

$$aX^i + \text{terms of higher degree}$$

lying in $\mathfrak{A}$. Then $\mathfrak{a}_i$ is an ideal of A, and $\mathfrak{a}_i \subset \mathfrak{a}_{i+1}$ (the proof of this assertion being the same as for polynomials). The ascending chain of ideals stops:

$$\mathfrak{a}_0 \subset \mathfrak{a}_1 \subset \mathfrak{a}_2 \subset \cdots \subset \mathfrak{a}_r = \mathfrak{a}_{r+1} = \cdots$$

As before, let a_{ij} ($i = 0, \ldots, r$ and $j = 1, \ldots, n_i$) be generators for the ideals $\mathfrak{a}_i$, and let f_{ij} be power series in A having a_{ij} as beginning coefficient. Given $f \in \mathfrak{A}$, starting with a term of degree d, say $d \leqq r$, we can find elements $c_1, \ldots, c_{nd} \in A$ such that

$$f - c_1 f_{d1} - \cdots - c_{n_d} f_{dn_d}$$

starts with a term of degree $\geqq d + 1$. Proceeding inductively, we may assume that $d > r$. We then use a linear combination

$$f - c_1^{(d)} X^{d-r} f_{r1} - \cdots - c_{n_r}^{(d)} X^{d-r} f_{rn_r}$$

to get a power series starting with a term of degree $\geqq d + 1$. In this way, if we start with a power series of degree $d > r$, then it can be expressed as a linear combination of $f_{r1}, \ldots, f_{rn_r}$ by means of the coefficients

$$g_1(X) = \sum_{\nu=d}^{\infty} c_1^{(\nu)} X^{\nu-r}, \ldots, g_{n_r}(X) = \sum_{\nu=d}^{\infty} c_{n_r}^{(\nu)} X^{\nu-r},$$

and we see that the f_{ij} generate our ideal $\mathfrak{A}$, as was to be shown.

Corollary 9.5. *If A is a Noetherian commutative ring, or a field, then $A[[X_1, \ldots, X_n]]$ is Noetherian.*

Examples. Power series in one variable are at the core of the theory of functions of one complex variable, and similarly for power series in several variables in the higher-dimensional case. See for instance [Gu 90].

Weierstrass polynomials occur in several contexts. First, they can be used to reduce questions about power series to questions about polynomials, in studying analytic sets. See for instance [GrH 78], Chapter 0. In a number-

theoretic context, such polynomials occur as characteristic polynomials in the Iwasawa theory of cyclotomic fields. Cf. [La 90], starting with Chapter 5.

Power series can also be used as generating functions. Suppose that to each positive integer n we associate a number $a(n)$. Then the **generating function** is the power series $\sum a(n)t^n$. In significant cases, it turns out that this function represents a rational function, and it may be a major result to prove that this is so.

For instance in Chapter X, §6 we shall consider a Poincaré series, associated with the length of modules. Similarly, in topology, consider a topological space X such that its homology groups (say) are finite dimensional over a field k of coefficients. Let $h_n = \dim H_n(X, k)$, where H_n is the n-th homology group. The **Poincaré series** is defined to be the generating series

$$P_X(t) = \sum h_n t^n.$$

Examples arise in the theory of dynamical systems. One considers a mapping $T: X \to X$ from a space X into itself, and we let N_n be the number of fixed points of the n-th iterate $T^n = T \circ T \circ \cdots \circ T$ (n times). The generating function is $\sum N_n t^n$. Because of the number of references I give here, I list them systematically at the end of the section. See first Artin–Mazur [ArM 65]; a proof by Manning of a conjecture of Smale [Ma 71]; and Shub's book [Sh 87], especially Chapter 10, Corollary 10.42 (Manning's theorem).

For an example in algebraic geometry, let V be an algebraic variety defined over a finite field k. Let K_n be the extension of k of degree n (in a given algebraic closure). Let N_n be the number of points of V in K_n. One defines the **zeta function** $Z(t)$ as the power series such that $Z(0) = 1$ and

$$Z'/Z(t) = \sum_{n=1}^{\infty} N_n t^{n-1}.$$

Then $Z(t)$ is a rational function (F. K. Schmidt when the dimension of V is 1, and Dwork in higher dimensions). For a discussion and references to the literature, see Appendix C of Hartshorne [Ha 77].

Finally we mention the **partition function** $p(n)$, which is the number of ways a positive integer can be expressed as a sum of positive integers. The generating function was determined by Euler to be

$$1 + \sum_{n=1}^{\infty} p(n)t^n = \prod_{n=1}^{\infty} (1 - t^n)^{-1}.$$

See for instance Hardy and Wright [HardW 71], Chapter XIX. The generating series for the partition function is related to the power series usually expressed in terms of a variable q, namely

$$\Delta = q \prod_{n=1}^{\infty} (1 - q^n)^{24} = \sum_{n=1}^{\infty} \tau(n) q^n,$$

which is the generating series for the **Ramanujan function** $\tau(n)$. The power series for Δ is also the expansion of a function in the theory of modular functions. For an introduction, see Serre's book [Se 73], last chapter, and books on elliptic functions, e.g. mine. We shall mention one application of the power series for Δ in the Galois theory chapter.

Generating power series also occur in K-theory, topological and algebraic geometric, as in Hirzebruch's formalism for the Riemann–Roch theorem and its extension by Grothendieck. See Atiyah [At 67], Hirzebruch [Hi 66], and [FuL 86]. I have extracted some formal elementary aspects having directly to do with power series in Exercises 21–27, which can be viewed as basic examples. See also Exercises 31–34 of the next chapter.

Bibliography

[ArM 65] M. ARTIN and B. MAZUR, On periodic points, *Ann. Math.* (2) **81** (1965) pp. 89–99

[At 67] M. ATIYAH, *K-Theory*, Addison-Wesley 1991 (reprinted from the Benjamin Lecture Notes, 1967)

[FuL 85] W. FULTON and S. LANG, *Riemann–Roch Algebra*, Springer-Verlag, New York, 1985

[GrH 78] P. GRIFFITHS and J. HARRIS, *Principles of Algebraic Geometry*, Wiley–Interscience, New York, 1978

[Gu 90] R. GUNNING, *Introduction to Holomorphic Functions of Several Variables*, Vol. II: *Local Theory*, Wadsworth and Brooks/Cole, 1990

[HardW 71] G. H. HARDY and E. M. WRIGHT, *An Introduction to the Theory of Numbers*, Oxford University Press, Oxford, UK, 1938–1971 (several editions)

[Hart 77] R. HARTSHORNE, *Algebraic Geometry*, Springer-Verlag, New York, 1977

[Hi 66] F. HIRZEBRUCH, *Topological Methods in Algebraic Geometry*, Springer-Verlag, New York, 1966 (translated and expanded from the original German, 1956)

[La 90] S. LANG, *Cyclotomic Fields*, I and II, Springer-Verlag, New York, 1990, combined edition of the original editions, 1978, 1980

[Ma 71] A. MANNING, Axiom A diffeomorphisms have rational zeta functions, *Bull. Lond. Math. Soc.* **3** (1971) pp. 215–220

[Se 73] J. P. SERRE, *A Course in Arithmetic*, Springer-Verlag, New York, 1973

[Sh 87] M. SHUB, *Global Stability of Dynamical Systems*, Springer-Verlag, New York, 1987

EXERCISES

1. Let k be a field and $f(X) \in k[X]$ a non-zero polynomial. Show that the following conditions are equivalent:
   (a) The ideal $(f(X))$ is prime.
   (b) The ideal $(f(X))$ is maximal.
   (c) $f(X)$ is irreducible.

2. (a) State and prove the analogue of Theorem 5.2 for the rational numbers.
   (b) State and prove the analogue of Theorem 5.3 for positive integers.

3. Let f be a polynomial in one variable over a field k. Let X, Y be two variables. Show that in $k[X, Y]$ we have a "Taylor series" expansion

$$f(X + Y) = f(X) + \sum_{i=1}^{n} \varphi_i(X) Y^i,$$

   where $\varphi_i(X)$ is a polynomial in X with coefficients in k. If k has characteristic 0, show that

$$\varphi_i(X) = \frac{D^i f(X)}{i!}.$$

4. Generalize the preceding exercise to polynomials in several variables (introduce partial derivatives and show that a finite Taylor expansion exists for a polynomial in several variables).

5. (a) Show that the polynomials $X^4 + 1$ and $X^6 + X^3 + 1$ are irreducible over the rational numbers.
   (b) Show that a polynomial of degree 3 over a field is either irreducible or has a root in the field. Is $X^3 - 5X^2 + 1$ irreducible over the rational numbers?
   (c) Show that the polynomial in two variables $X^2 + Y^2 - 1$ is irreducible over the rational numbers. Is it irreducible over the complex numbers?

6. Prove the integral root test of §3.

7. (a) Let k be a finite field with q elements. Let $f(X_1, \ldots, X_n)$ be a polynomial in $k[X]$ of degree d and assume $f(0, \ldots, 0) = 0$. An element $(a_1, \ldots, a_n) \in k^{(n)}$ such that $f(a) = 0$ is called a zero of f. If $n > d$, show that f has at least one other zero in $k^{(n)}$. [*Hint*: Assume the contrary, and compare the degrees of the reduced polynomial belonging to

$$1 - f(X)^{q-1}$$

   and $(1 - X_1^{q-1}) \cdots (1 - X_n^{q-1})$. The theorem is due to Chevalley.]
   (b) Refine the above results by proving that the number N of zeros of f in $k^{(n)}$ is $\equiv 0 \pmod p$, arguing as follows. Let i be an integer $\geqq 1$. Show that

$$\sum_{x \in k} x^i = \begin{cases} q - 1 = -1 & \text{if } q - 1 \text{ divides } i, \\ 0 & \text{otherwise.} \end{cases}$$

   Denote the preceding function of i by $\psi(i)$. Show that

$$N \equiv \sum_{x \in k^{(n)}} (1 - f(x)^{q-1})$$

and for each n-tuple $(i_1, \ldots, i_n)$ of integers $\geqq 0$ that

$$\sum_{x \in k^{(n)}} x_1^{i_1} \cdots x_n^{i_n} = \psi(i_1) \cdots \psi(i_n).$$

Show that both terms in the sum for N above yield $0 \bmod p$. (The above argument is due to Warning.)

(c) Extend Chevalley's theorem to r polynomials $f_1, \ldots, f_r$ of degrees $d_1, \ldots, d_r$ respectively, in n variables. If they have no constant term and $n > \sum d_i$, show that they have a non-trivial common zero.

(d) Show that an arbitrary function $f: k^{(n)} \to k$ can be represented by a polynomial. (As before, k is a finite field.)

8. Let A be a commutative entire ring and X a variable over A. Let $a, b \in A$ and assume that a is a unit in A. Show that the map $X \mapsto aX + b$ extends to a unique automorphism of $A[X]$ inducing the identity on A. What is the inverse automorphism?

9. Show that every automorphism of $A[X]$ is of the type described in Exercise 8.

10. Let K be a field, and $K(X)$ the quotient field of $K[X]$. Show that every automorphism of $K(X)$ which induces the identity on K is of type

$$X \mapsto \frac{aX + b}{cX + d}$$

with $a, b, c, d \in K$ such that $(aX + b)/(cX + d)$ is not an element of K, or equivalently, $ad - bc \neq 0$.

11. Let A be a commutative entire ring and let K be its quotient field. We show here that some formulas from calculus have a purely algebraic setting. Let $D: A \to A$ be a **derivation**, that is an additive homomorphism satisfying the rule for the derivative of a product, namely

$$D(xy) = xDy + yDx \qquad \text{for} \quad x, y \in A.$$

(a) Prove that D has a unique extension to a derivation of K into itself, and that this extension satisfies the rule

$$D(x/y) = \frac{yDx - xDy}{y^2}$$

for $x, y \in A$ and $y \neq 0$. [Define the extension by this formula, prove that it is independent of the choice of x, y to write the fraction x/y, and show that it is a derivation having the original value on elements of A.]

(b) Let $L(x) = Dx/x$ for $x \in K^*$. Show that $L(xy) = L(x) + L(y)$. The homomorphism L is called the **logarithmic derivative**.

(c) Let D be the standard derivative in the polynomial ring $k[X]$ over a field k. Let $R(X) = c \prod (X - \alpha_i)^{m_i}$ with $\alpha_i \in k$, $c \in k$, and $m_i \in \mathbf{Z}$, so $R(X)$ is a rational

function. Show that

$$R'/R = \sum \frac{m_i}{X - \alpha_i}.$$

12. (a) If $f(X) = aX^2 + bX + c$, show that the discriminant of f is $b^2 - 4ac$.
    (b) If $f(X) = a_0X^3 + a_1X^2 + a_2X + a_3$, show that the discriminant of f is

    $$a_1^2a_2^2 - 4a_0a_2^3 - 4a_1^3a_3 - 27a_0^2a_3^2 + 18a_0a_1a_2a_3.$$

    (c) Let $f(X) = (X - t_1) \cdots (X - t_n)$. Show that

    $$D_f = (-1)^{n(n-1)/2} \prod_{i=1}^{n} f'(t_i).$$

13. Polynomials will be taken over an algebraically closed field of characteristic 0.
    (a) Prove

    Davenport's theorem. *Let $f(t)$, $g(t)$ be polynomials such that $f^3 - g^2 \neq 0$. Then*

    $$\deg(f^3 - g^2) \geqq \tfrac{1}{2} \deg f + 1.$$

    Or put another way, let $h = f^3 - g^2$ and assume $h \neq 0$. Then

    $$\deg f \leqq 2 \deg h - 2.$$

    To do this, first assume f, g relatively prime and apply Mason's theorem. In general, proceed as follows.
    (b) *Let A, B, f, g be polynomials such that Af, Bg are relatively prime $\neq 0$. Let $h = Af^3 + Bg^2$. Then*

    $$\deg f \leqq \deg A + \deg B + 2 \deg h - 2.$$

    This follows directly from Mason's theorem. Then starting with f, g not necessarily relatively prime, start factoring out common factors until no longer possible, to effect the desired reduction. When I did it, I needed to do this step three times, so don't stop until you get it.
    (c) Generalize (b) to the case of $f^m - g^n$ for arbitrary positive integer exponents m and n.

14. Prove that the generalized Szpiro conjecture implies the *abc* conjecture.

15. Prove that the *abc* conjecture implies the following conjecture: There are infinitely many primes p such that $2^{p-1} \not\equiv 1 \bmod p^2$. [Cf. the reference [Sil 88] and [La 90] at the end of §7.]

16. Let w be a complex number, and let $c = \max(1, |w|)$. Let F, G be non-zero polynomials in one variable with complex coefficients, of degrees d and d' respectively, such that $|F|, |G| \geqq 1$. Let R be their resultant. Then

    $$|R| \leqq c^{d+d'}[|F(w)| + |G(w)|]|F|^{d'}|G|^{d}(d + d')^{d+d'}.$$

    (We denote by $|F|$ the maximum of the absolute values of the coefficients of F.)

17. Let d be an integer $\geqq 3$. Prove the existence of an irreducible polynomial of degree d over $\mathbf{Q}$, having precisely $d - 2$ real roots, and a pair of complex conjugate roots. Use the following construction. Let $b_1, \ldots, b_{d-2}$ be distinct

integers, and let a be an integer > 0. Let

$$g(X) = (X^2 + a)(X - b_1) \cdots (X - b_{d-1}) = X^d + c_{d-1}X^{d-1} + \cdots + c_0.$$

Observe that $c_i \in \mathbf{Z}$ for all i. Let p be a prime number, and let

$$g_n(X) = g(X) + \frac{p}{p^{dn}}$$

so that g_n converges to g (i.e. the coefficients of g_n converge to the coefficients of g).

(a) Prove that g_n has precisely $d - 2$ real roots for n sufficiently large. (You may use a bit of calculus, or use whatever method you want.)

(b) Prove that g_n is irreducible over $\mathbf{Q}$.

Integral-valued polynomials

18. Let $P(X) \in \mathbf{Q}[X]$ be a polynomial in one variable with rational coefficients. It may happen that $P(n) \in \mathbf{Z}$ for all sufficiently large integers n without necessarily P having integer coefficients.

(a) Give an example of this.

(b) Assume that P has the above property. Prove that there are integers $c_0, c_1, \ldots, c_r$ such that

$$P(X) = c_0\binom{X}{r} + c_1\binom{X}{r-1} + \cdots + c_r,$$

where

$$\binom{X}{r} = \frac{1}{r!}X(X-1)\cdots(X-r+1)$$

is the binomial coefficient function. In particular, $P(n) \in \mathbf{Z}$ for all n. Thus we may call P **integral valued**.

(c) Let $f: \mathbf{Z} \to \mathbf{Z}$ be a function. Assume that there exists an integral valued polynomial Q such that the difference function Δf defined by

$$(\Delta f)(n) = f(n) - f(n-1)$$

is equal to $Q(n)$ for all n sufficiently large. Show that there exists an integral-valued polynomial P such that $f(n) = P(n)$ for all n sufficiently large.

Exercises on symmetric functions

19. (a) Let $X_1, \ldots, X_n$ be variables. Show that any homogeneous polynomial in $\mathbf{Z}[X_1, \ldots, X_n]$ of degree $> n(n-1)$ lies in the ideal generated by the elementary symmetric functions $s_1, \ldots, s_n$.

(b) With the same notation show that $\mathbf{Z}[X_1, \ldots, X_n]$ is a free $\mathbf{Z}[s_1, \ldots, s_n]$ module with basis the monomials

$$X^{(r)} = X_1^{r_1} \cdots X_n^{r_n}$$

with $0 \leqq r_i \leqq n - i$.

(c) Let $X_1, \dots, X_n$ and $Y_1, \dots, Y_m$ be two independent sets of variables. Let $s_1, \dots, s_n$ be the elementary symmetric functions of X and $s'_1, \dots, s'_m$ the elementary symmetric functions of Y (using vector vector notation). Show that $\mathbf{Z}[X, Y]$ is free over $\mathbf{Z}[s, s']$ with basis $X^{(r)}Y^{(q)}$, and the exponents (r), (q) satisfying inequalities as in (b).

(d) Let I be an ideal in $\mathbf{Z}[s, s']$. Let J be the ideal generated by I in $\mathbf{Z}[X, Y]$. Show that

$$J \cap \mathbf{Z}[s, s'] = I.$$

20. Let A be a commutative ring. Let t be a variable. Let

$$f(t) = \sum_{i=0}^{m} a_i t^i \qquad \text{and} \qquad g(t) = \sum_{i=0}^{n} b_i t^i$$

be polynomials whose constant terms are $a_0 = b_0 = 1$. If

$$f(t)g(t) = 1,$$

show that there exists an integer N $\big(= (m + n)(m + n - 1)\big)$ such that any monomial

$$a_1^{r_1} \cdots a_n^{r_n}$$

with $\sum jr_j > N$ is equal to 0. [*Hint*: Replace the a's and b's by variables. Use Exercise 19(b) to show that any monomial $M(a)$ of weight $> N$ lies in the ideal I generated by the elements

$$c_k = \sum_{i=0}^{k} a_i b_{k-i}$$

(letting $a_0 = b_0 = 1$). Note that c_k is the k-th elementary symmetric function of the $m + n$ variables (X, Y).]

[*Note*: For some interesting contexts involving symmetric functions, see Cartier's talk at the Bourbaki Seminar, 1982–1983.]

λ-rings

The following exercises start a train of thought which will be pursued in Exercise 33 of Chapter V; Exercises 22–24 of Chapter XVIII; and Chapter XX, §3. These originated to a large extent in Hirzebruch's Riemann–Roch theorem and its extension by Grothendieck who defined λ-rings in general.

Let K be a commutative ring. By **λ-operations** we mean a family of mappings

$$\lambda^i : K \to K$$

for each integer $i \geqq 0$ satisfying the relations for all $x \in K$:

$$\lambda^0(x) = 1, \qquad \lambda^1(x) = x,$$

and for all integers $n \geqq 0$, and $x, y \in K$,

$$\lambda^n(x + y) = \sum_{i=0}^{n} \lambda^i(x)\lambda^{n-i}(y).$$

The reader will meet examples of such operations in the chapter on the alternating and symmetric products, but the formalism of such operations depends only on the above relations, and so can be developed here in the context of formal power series. Given a λ-operation, in which case we also say that K is a **λ-ring,** we define the power series

$$\lambda_t(x) = \sum_{i=0}^{\infty} \lambda^i(x)t^i.$$

Prove the following statements.

21. The map $x \mapsto \lambda_t(x)$ is a homomorphism from the additive group of K into the multiplicative group of power series $1 + tK[[t]]$ whose constant term is equal to 1. Conversely, any such homomorphism such that $\lambda_t(x) = 1 + xt +$ higher terms gives rise to λ-operations.

22. Let $s = at +$ higher terms be a power series in $K[[t]]$ such that a is a unit in K. Show that there is a power series

$$t = g(s) = \sum b_i s^i \qquad \text{with} \quad b_i \in K.$$

Show that any power series $f(t) \in K[[t]]$ can be written in the form $h(s)$ for some other power series with coefficients in K.

Given a λ-operation on K, define the corresponding **Grothendieck power series**

$$\gamma_t(x) = \lambda_{t/(1-t)}(x) = \lambda_s(x)$$

where $s = t/(1 - t)$. Then the map

$$x \mapsto \gamma_t(x)$$

is a homomorphism as before. We define $\gamma^i(x)$ by the relation

$$\gamma_t(x) = \sum \gamma^i(x)t^i.$$

Show that γ satisfies the following properties.

23. (a) For every integer $n \geqq 0$ we have

$$\gamma^n(x + y) = \sum_{i=0}^{n} \gamma^i(x)\gamma^{n-i}(y).$$

(b) $\gamma_t(1) = 1/(1 - t)$.
(c) $\gamma_t(-1) = 1 - t$.

24. Assume that $\lambda^i u = 0$ for $i > 1$. Show:
(a) $\gamma_t(u - 1) = 1 + (u - 1)t$.
(b) $\gamma_t(1 - u) = \sum_{i=0}^{\infty} (1 - u)^i t^i$.

25. **Bernoulli numbers.** Define the Bernoulli numbers B_k as the coefficients in the power series

$$F(t) = \frac{t}{e^t - 1} = \sum_{k=0}^{\infty} B_k \frac{t^k}{k!}.$$

Of course, $e^t = \sum t^n/n!$ is the standard power series with rational coefficients $1/n!$. Prove:

(a) $B_0 = 1$, $B_1 = -\frac{1}{2}$, $B_2 = \frac{1}{6}$.

(b) $F(-t) = t + F(t)$, and $B_k = 0$ if k is odd $\neq 1$.

26. **Bernoulli polynomials.** Define the Bernoulli polynomials $\mathbf{B}_k(X)$ by the power series expansion

$$F(t, X) = \frac{te^{tX}}{e^t - 1} = \sum_{k=0}^{\infty} \mathbf{B}_k(X)\frac{t^k}{k!}.$$

It is clear that $B_k = \mathbf{B}_k(0)$, so the Bernoulli numbers are the constant terms of the Bernoulli polynomials. Prove:

(a) $\mathbf{B}_0(X) = 1$, $\mathbf{B}_1(X) = X - \frac{1}{2}$, $\mathbf{B}_2(X) = X^2 - X + \frac{1}{6}$.

(b) For each positive integer N,

$$\mathbf{B}_k(X) = N^{k-1} \sum_{a=0}^{N-1} \mathbf{B}_k\left(\frac{X + a}{N}\right).$$

(c) $\mathbf{B}_k(X) = X^k - \frac{1}{2}kX^{k-1} +$ lower terms.

(d) $F(t, X + 1) - F(t, X) = te^{Xt} = t\sum X^k \dfrac{t^k}{k!}$.

(e) $\mathbf{B}_k(X + 1) - \mathbf{B}_k(X) = kX^{k-1}$ for $k \geqq 1$.

27. Let N be a positive integer and let f be a function on $\mathbf{Z}/N\mathbf{Z}$. Form the power series

$$F_f(t, X) = \sum_{a=0}^{N-1} f(a)\frac{te^{(a+X)t}}{e^{Nt} - 1}.$$

Following Leopoldt, define the **generalized Bernoulli polynomials** relative to the function f by

$$F_f(t, X) = \sum_{k=0}^{\infty} \mathbf{B}_{k,f}(X)\frac{t^k}{k!}.$$

In particular, the constant term of $\mathbf{B}_{k,f}(X)$ is defined to be the **generalized Bernoulli number** $B_{k,f} = \mathbf{B}_{k,f}(0)$ introduced by Leopoldt in cyclotomic fields. Prove:

(a) $F_f(t, X + k) = e^{kt}F_f(t, X)$.

(b) $F_f(t, X + N) - F_f(t, X) = (e^{Nt} - 1)F_f(t, X)$.

(c) $\dfrac{1}{k}[\mathbf{B}_{k,f}(X + N) - \mathbf{B}_{k,f}(X)] = \sum_{a=0}^{N-1} f(a)(a + X)^{k-1}$.

(d)
$$\mathbf{B}_{k,f}(X) = \sum_{i=0}^{k} \binom{k}{i} B_{i,f}X^{n-i}$$
$$= B_{k,f} + kB_{k-1,f}X + \cdots + kB_{1,f}X^{k-1} + B_{0,f}X^k.$$

Note. The exercises on Bernoulli numbers and polynomials are designed not only to give examples for the material in the text, but to show how this material leads into major areas of mathematics: in topology and algebraic geometry centering

around Riemann–Roch theorems; analytic and algebraic number theory, as in the theory of the zeta functions and the theory of modular forms, cf. my *Introduction to Modular Forms*, Springer-Verlag, New York, 1976, Chapters XIV and XV; my *Cyclotomic Fields*, I and II, Springer-Verlag, New York, 1990, Chapter 2, §2; Kubert–Lang's *Modular Units*, Springer-Verlag, New York, 1981; etc.

Further Comments, 1996–2001. I was informed by Umberto Zannier that what has been called Mason's theorem was proved three years earlier by Stothers [Sto 81], Theorem 1.1. Zannier himself has published some results on Davenport's theorem [Za 95], without knowing of the paper by Stothers, using a method similar to that of Stothers, and rediscovering some of Stothers' results, but also going beyond. Indeed, Stothers uses the "Belyi method" belonging to algebraic geometry, and increasingly appearing as a fundamental tool. Mason gave a very elementary proof, accessible at the basic level of algebra. An even shorter and very elegant proof of the Mason-Stothers theorem was given by Noah Snyder [Sny 00]. I am much indebted to Snyder for showing me that proof before publication, and I reproduced it in [La 99b]. But I recommend looking at Snyder's version.

[La 99b] S. LANG, *Math Talks for Undergraduates*, Springer Verlag 1999

[Sny 00] N. SNYDER, An alternate proof of Mason's theorem, *Elemente der Math.* **55** (2000) pp. 93–94

[Sto 81] W. STOTHERS, Polynomial identities and hauptmoduln, *Quart. J. Math. Oxford* (2) **32** (1981) pp. 349–370

[Za 95] U. ZANNIER, On Davenport's bound for the degree of $f^3 - g^2$ and Riemann's existence theorem, *Acta Arithm.* **LXXI.2** (1995) pp. 107–137

Part Two

ALGEBRAIC EQUATIONS

This part is concerned with the solutions of algebraic equations, in one or several variables. This is the recurrent theme in every chapter of this part, and we lay the foundations for all further studies concerning such equations.

Given a subring A of a ring B, and a finite number of polynomials $f_1, \ldots, f_n$ in $A[X_1, \ldots, X_n]$, we are concerned with the n-tuples

$$(b_1, \ldots, b_n) \in B^{(n)}$$

such that

$$f_i(b_1, \ldots, b_n) = 0$$

for $i = 1, \ldots, r$. For suitable choices of A and B, this includes the general problem of diophantine analysis when A, B have an "arithmetic" structure.

We shall study various cases. We begin by studying roots of one polynomial in one variable over a field. We prove the existence of an algebraic closure, and emphasize the role of irreducibility.

Next we study the group of automorphisms of algebraic extensions of a field, both intrinsically and as a group of permutations of the roots of a polynomial. We shall mention some major unsolved problems along the way.

It is also necessary to discuss extensions of a ring, to give the possibility of analyzing families of extensions. The ground work is laid in Chapter VII.

In Chapter IX, we come to the zeros of polynomials in several variables, essentially over algebraically closed fields. But again, it is advantageous to

consider polynomials over rings, especially **Z**, since in projective space, the conditions that homogeneous polynomials have a non-trivial common zero can be given universally over **Z** in terms of their coefficients.

Finally we impose additional structures like those of reality, or metric structures given by absolute values. Each one of these structures gives rise to certain theorems describing the structure of the solutions of equations as above, and especially proving the existence of solutions in important cases.

CHAPTER V

Algebraic Extensions

In this first chapter concerning polynomial equations, we show that given a polynomial over a field, there always exists some extension of the field where the polynomial has a root, and we prove the existence of an algebraic closure. We make a preliminary study of such extensions, including the automorphisms, and we give algebraic extensions of finite fields as examples.

§1. FINITE AND ALGEBRAIC EXTENSIONS

Let F be a field. If F is a subfield of a field E, then we also say that E is an **extension field** of F. We may view E as a vector space over F, and we say that E is a **finite** or **infinite** extension of F according as the dimension of this vector space is finite or infinite.

Let F be a subfield of a field E. An element α of E is said to be **algebraic** over F if there exist elements $a_0, \ldots, a_n$ $(n \geqq 1)$ of F, not all equal to 0, such that

$$a_0 + a_1\alpha + \cdots + a_n\alpha^n = 0.$$

If $\alpha \neq 0$, and α is algebraic, then we can always find elements a_i as above such that $a_0 \neq 0$ (factoring out a suitable power of α).

Let X be a variable over F. We can also say that α is algebraic over F if the homomorphism

$$F[X] \to E$$

which is the identity on F and maps X on α has a non-zero kernel. In that case the kernel is an ideal which is principal, generated by a single polynomial $p(X)$, which we may assume has leading coefficient 1. We then have an isomorphism

$$F[X]/(p(X)) \approx F[\alpha],$$

and since $F[\alpha]$ is entire, it follows that $p(X)$ is irreducible. Having normalized $p(X)$ so that its leading coefficient is 1, we see that $p(X)$ is uniquely determined by α and will be called THE irreducible polynomial of α over F. We sometimes denote it by $\operatorname{Irr}(\alpha, F, X)$.

An extension E of F is said to be **algebraic** if every element of E is algebraic over F.

Proposition 1.1. *Let E be a finite extension of F. Then E is algebraic over F.*

Proof. Let $\alpha \in E$, $\alpha \neq 0$. The powers of α,

$$1, \alpha, \alpha^2, \dots, \alpha^n,$$

cannot be linearly independent over F for all positive integers n, otherwise the dimension of E over F would be infinite. A linear relation between these powers shows that α is algebraic over F.

Note that the converse of Proposition 1.1 is not true; there exist infinite algebraic extensions. We shall see later that the subfield of the complex numbers consisting of all algebraic numbers over $\mathbf{Q}$ is an infinite extension of $\mathbf{Q}$.

If E is an extension of F, we denote by

$$[E : F]$$

the dimension of E as vector space over F. It may be infinite.

Proposition 1.2. *Let k be a field and $F \subset E$ extension fields of k. Then*

$$[E : k] = [E : F][F : k].$$

If $\{x_i\}_{i \in I}$ is a basis for F over k and $\{y_j\}_{j \in J}$ is a basis for E over F, then $\{x_i y_j\}_{(i,j) \in I \times J}$ is a basis for E over k.

Proof. Let $z \in E$. By hypothesis there exist elements $\alpha_j \in F$, almost all $\alpha_j = 0$, such that

$$z = \sum_{j \in J} \alpha_j y_j.$$

For each $j \in J$ there exist elements $b_{ji} \in k$, almost all of which are equal to 0, such that

$$\alpha_j = \sum_{i \in I} b_{ji} x_i,$$

and hence

$$z = \sum_j \sum_i b_{ji} x_i y_j.$$

This shows that $\{x_i y_j\}$ is a family of generators for E over k. We must show that it is linearly independent. Let $\{c_{ij}\}$ be a family of elements of k, almost all of which are 0, such that

$$\sum_j \sum_i c_{ij} x_i y_j = 0.$$

Then for each j,

$$\sum_i c_{ij} x_i = 0$$

because the elements y_j are linearly independent over F. Finally $c_{ij} = 0$ for each i because $\{x_i\}$ is a basis of F over k, thereby proving our proposition.

Corollary 1.3. *The extension E of k is finite if and only if E is finite over F and F is finite over k.*

As with groups, we define a **tower** of fields to be a sequence

$$F_1 \subset F_2 \subset \cdots \subset F_n$$

of extension fields. The tower is called **finite** if and only if each step is finite.

Let k be a field, E an extension field, and $\alpha \in E$. We denote by $k(\alpha)$ the smallest subfield of E containing both k and α. It consists of all quotients $f(\alpha)/g(\alpha)$, where f, g are polynomials with coefficients in k and $g(\alpha) \neq 0$.

Proposition 1.4. *Let α be algebraic over k. Then $k(\alpha) = k[\alpha]$, and $k(\alpha)$ is finite over k. The degree $[k(\alpha) : k]$ is equal to the degree of* $\mathrm{Irr}(\alpha, k, X)$.

Proof. Let $p(X) = \mathrm{Irr}(\alpha, k, X)$. Let $f(X) \in k[X]$ be such that $f(\alpha) \neq 0$. Then $p(X)$ does not divide $f(X)$, and hence there exist polynomials $g(X)$, $h(X) \in k[X]$ such that

$$g(X)p(X) + h(X)f(X) = 1.$$

From this we get $h(\alpha)f(\alpha) = 1$, and we see that $f(\alpha)$ is invertible in $k[\alpha]$. Hence $k[\alpha]$ is not only a ring but a field, and must therefore be equal to $k(\alpha)$. Let $d = \deg p(X)$. The powers

$$1, \alpha, \ldots, \alpha^{d-1}$$

are linearly independent over k, for otherwise suppose

$$a_0 + a_1 \alpha + \cdots + a_{d-1} \alpha^{d-1} = 0$$

with $a_i \in k$, not all $a_i = 0$. Let $g(X) = a_0 + \cdots + a_{d-1}X^{d-1}$. Then $g \neq 0$ and $g(\alpha) = 0$. Hence $p(X)$ divides $g(X)$, contradiction. Finally, let $f(\alpha) \in k[\alpha]$, where $f(X) \in k[X]$. There exist polynomials $q(X)$, $r(X) \in k[X]$ such that $\deg r < d$ and

$$f(X) = q(X)p(X) + r(X).$$

Then $f(\alpha) = r(\alpha)$, and we see that $1, \alpha, \ldots, \alpha^{d-1}$ generate $k[\alpha]$ as a vector space over k. This proves our proposition.

Let E, F be extensions of a field k. If E and F are contained in some field L then we denote by EF the smallest subfield of L containing both E and F, and call it the **compositum** of E and F, in L. If E, F are not given as embedded in a common field L, then we cannot define the compositum.

Let k be a subfield of E and let $\alpha_1, \ldots, \alpha_n$ be elements of E. We denote by

$$k(\alpha_1, \ldots, \alpha_n)$$

the smallest subfield of E containing k and $\alpha_1, \ldots, \alpha_n$. Its elements consist of all quotients

$$\frac{f(\alpha_1, \ldots, \alpha_n)}{g(\alpha_1, \ldots, \alpha_n)}$$

where f, g are polynomials in n variables with coefficients in k, and

$$g(\alpha_1, \ldots, \alpha_n) \neq 0.$$

Indeed, the set of such quotients forms a field containing k and $\alpha_1, \ldots, \alpha_n$. Conversely, any field containing k and

$$\alpha_1, \ldots, \alpha_n$$

must contain these quotients.

We observe that E is the union of all its subfields $k(\alpha_1, \ldots, \alpha_n)$ as $(\alpha_1, \ldots, \alpha_n)$ ranges over finite subfamilies of elements of E. We could define the *compositum of an arbitrary subfamily of subfields of a field L* as the smallest subfield containing all fields in the family. We say that E is **finitely generated** over k if there is a finite family of elements $\alpha_1, \ldots, \alpha_n$ of E such that

$$E = k(\alpha_1, \ldots, \alpha_n).$$

We see that E is the compositum of all its finitely generated subfields over k.

Proposition 1.5. *Let E be a finite extension of k. Then E is finitely generated.*

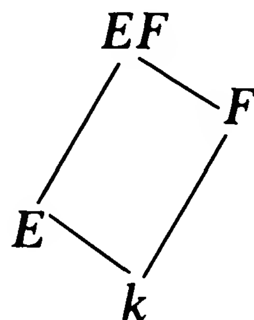
Proof. Let $\{\alpha_1, \ldots, \alpha_n\}$ be a basis of E as vector space over k. Then certainly

$$E = k(\alpha_1, \ldots, \alpha_n).$$

If $E = k(\alpha_1, \ldots, \alpha_n)$ is finitely generated, and F is an extension of k, both F, E contained in L, then

$$EF = F(\alpha_1, \ldots, \alpha_n),$$

and EF is finitely generated over F. We often draw the following picture:



Lines slanting up indicate an inclusion relation between fields. We also call the extension EF of F the **translation** of E to F, or also the **lifting** of E to F.

Let α be algebraic over the field k. Let F be an extension of k, and assume $k(\alpha)$, F both contained in some field L. Then α is algebraic over F. Indeed, the irreducible polynomial for α over k has *a fortiori* coefficients in F, and gives a linear relation for the powers of α over F.

Suppose that we have a tower of fields:

$$k \subset k(\alpha_1) \subset k(\alpha_1, \alpha_2) \subset \cdots \subset k(\alpha_1, \ldots, \alpha_n),$$

each one generated from the preceding field by a single element. Assume that each α_i is algebraic over k, $i = 1, \ldots, n$. As a special case of our preceding remark, we note that α_{i+1} is algebraic over $k(\alpha_1, \ldots, \alpha_i)$. Hence each step of the tower is algebraic.

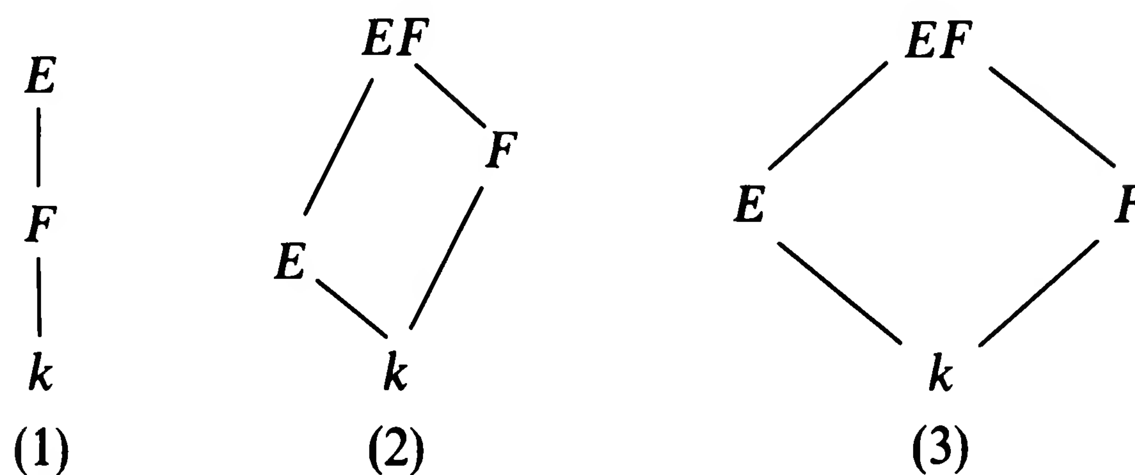
Proposition 1.6. *Let $E = k(\alpha_1, \ldots, \alpha_n)$ be a finitely generated extension of a field k, and assume α_i algebraic over k for each $i = 1, \ldots, n$. Then E is finite algebraic over k.*

Proof. From the above remarks, we know that E can be obtained as the end of a tower each of whose steps is generated by one algebraic element, and is therefore finite by Proposition 1.4. We conclude that E is finite over k by Corollary 1.3, and that it is algebraic by Proposition 1.1.

Let $\mathcal{C}$ be a certain class of extension fields $F \subset E$. We shall say that $\mathcal{C}$ is **distinguished** if it satisfies the following conditions:

(1) Let $k \subset F \subset E$ be a tower of fields. The extension $k \subset E$ is in $\mathcal{C}$ if and only if $k \subset F$ is in $\mathcal{C}$ and $F \subset E$ is in $\mathcal{C}$.
(2) If $k \subset E$ is in $\mathcal{C}$, if F is any extension of k, and E, F are both contained in some field, then $F \subset EF$ is in $\mathcal{C}$.
(3) If $k \subset F$ and $k \subset E$ are in $\mathcal{C}$ and F, E are subfields of a common field, then $k \subset FE$ is in $\mathcal{C}$.

The diagrams illustrating our properties are as follows:



These lattice diagrams of fields are extremely suggestive in handling extension fields.

We observe that (3) follows formally from the first two conditions. Indeed, one views EF over k as a tower with steps $k \subset F \subset EF$.

As a matter of notation, it is convenient to write E/F instead of $F \subset E$ to denote an extension. There can be no confusion with factor groups since we shall never use the notation E/F to denote such a factor group when E is an extension field of F.

Proposition 1.7. *The class of algebraic extensions is distinguished, and so is the class of finite extensions.*

Proof. Consider first the class of finite extensions. We have already proved condition (1). As for (2), assume that E/k is finite, and let F be any extension of k. By Proposition 1.5 there exist elements $\alpha_1, \dots, \alpha_n \in E$ such that $E = k(\alpha_1, \dots, \alpha_n)$. Then $EF = F(\alpha_1, \dots, \alpha_n)$, and hence EF/F is finitely generated by algebraic elements. Using Proposition 1.6 we conclude that EF/F is finite.

Consider next the class of algebraic extensions, and let

$$k \subset F \subset E$$

be a tower. Assume that E is algebraic over k. Then *a fortiori*, F is algebraic over k and E is algebraic over F. Conversely, assume each step in the tower to be algebraic. Let $\alpha \in E$. Then α satisfies an equation

$$a_n\alpha^n + \cdots + a_0 = 0$$

with $a_i \in F$, not all $a_i = 0$. Let $F_0 = k(a_n, \dots, a_0)$. Then F_0 is finite over k by Proposition 1.6, and α is algebraic over F_0. From the tower

$$k \subset F_0 = k(a_n, \dots, a_0) \subset F_0(\alpha)$$

and the fact that each step in this tower is finite, we conclude that $F_0(\alpha)$ is finite over k, whence α is algebraic over k, thereby proving that E is algebraic over k and proving condition (1) for algebraic extensions. Condition (2) has already been observed to hold, i.e. an element remains algebraic under lifting, and hence so does an extension.

Remark. It is true that finitely generated extensions form a distinguished class, but one argument needed to prove part of (1) can be carried out only with more machinery than we have at present. Cf. the chapter on transcendental extensions.

§2. ALGEBRAIC CLOSURE

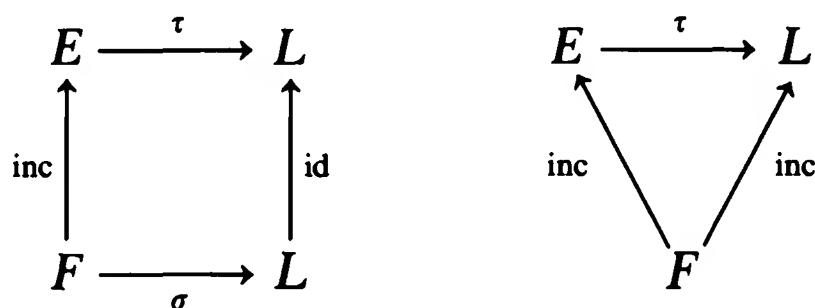
In this and the next section we shall deal with embeddings of a field into another. We therefore define some terminology.

Let E be an extension of a field F and let

$$\sigma: F \to L$$

be an embedding (i.e. an injective homomorphism) of F into L. Then σ induces an isomorphism of F with its image σF, which is sometimes written F^σ. An embedding τ of E in L will be said to be **over** σ if the restriction of τ to F is equal to σ. We also say that τ **extends** σ. If σ is the identity then we say that τ is an embedding of E **over** F.

These definitions could be made in more general categories, since they depend only on diagrams to make sense:



Remark. Let $f(X) \in F[X]$ be a polynomial, and let α be a root of f in E. Say $f(X) = a_0 + \cdots + a_n X^n$. Then

$$0 = f(\alpha) = a_0 + a_1\alpha + \cdots + a_n\alpha^n.$$

If τ extends σ as above, then we see that $\tau\alpha$ is a root of f^σ because

$$0 = \tau(f(\alpha)) = a_0^\sigma + a_1^\sigma(\tau\alpha) + \cdots + a_n^\sigma(\tau\alpha)^n.$$

Here we have written a^σ instead of $\sigma(a)$. This exponential notation is frequently convenient and will be used again in the sequel. Similarly, we write F^σ instead of $\sigma(F)$ or σF.

In our study of embeddings it will also be useful to have a lemma concerning embeddings of algebraic extensions into themselves. For this we note that if $\sigma: E \to L$ is an embedding over k (i.e. inducing the identity on k), then σ can be viewed as a k-homomorphism of vector spaces, because both E, L can be viewed as vector spaces over k. Furthermore σ is injective.

Lemma 2.1. *Let E be an algebraic extension of k, and let $\sigma: E \to E$ be an embedding of E into itself over k. Then σ is an automorphism.*

Proof. Since σ is injective, it will suffice to prove that σ is surjective. Let α be an element of E, let $p(X)$ be its irreducible polynomial over k, and let E' be the subfield of E generated by all the roots of $p(X)$ which lie in E. Then E' is finitely generated, hence is a finite extension of k. Furthermore, σ must map a root of $p(X)$ on a root of $p(X)$, and hence σ maps E' into itself. We can view σ as a k-homomorphism of vector spaces because σ induces the identity on k. Since σ is injective, its image $\sigma(E')$ is a subspace of E' having the same dimension $[E' : k]$. Hence $\sigma(E') = E'$. Since $\alpha \in E'$, it follows that α is in the image of σ, and our lemma is proved.

Let E, F be extensions of a field k, contained in some bigger field L. We can form the ring $E[F]$ generated by the elements of F over E. Then $E[F] = F[E]$, and EF is the quotient field of this ring. It is clear that the elements of $E[F]$ can be written in the form

$$a_1 b_1 + \cdots + a_n b_n$$

with $a_i \in E$ and $b_i \in F$. Hence EF is the field of quotients of these elements.

Lemma 2.2. *Let E_1, E_2 be extensions of a field k, contained in some bigger field E, and let σ be an embedding of E in some field L. Then*

$$\sigma(E_1 E_2) = \sigma(E_1)\sigma(E_2).$$

Proof. We apply σ to a quotient of elements of the above type, say

$$\sigma\left(\frac{a_1 b_1 + \cdots + a_n b_n}{a_1' b_1' + \cdots + a_m' b_m'}\right) = \frac{a_1^\sigma b_1^\sigma + \cdots + a_n^\sigma b_n^\sigma}{a_1'^\sigma b_1'^\sigma + \cdots + a_m'^\sigma b_m'^\sigma},$$

and see that the image is an element of $\sigma(E_1)\sigma(E_2)$. It is clear that the image $\sigma(E_1 E_2)$ is $\sigma(E_1)\sigma(E_2)$.

Let k be a field, $f(X)$ a polynomial of degree $\geqq 1$ in $k[X]$. We consider the problem of finding an extension E of k in which f has a root. If $p(X)$ is an irreducible polynomial in $k[X]$ which divides $f(X)$, then any root of $p(X)$ will also be a root of $f(X)$, so we may restrict ourselves to irreducible polynomials.

Let $p(X)$ be irreducible, and consider the canonical homomorphism

$$\sigma: k[X] \to k[X]/(p(X)).$$

Then σ induces a homomorphism on k, whose kernel is 0, because every nonzero element of k is invertible in k, generates the unit ideal, and 1 does not lie in the kernel. Let ξ be the image of X under σ, i.e. $\xi = \sigma(X)$ is the residue class of X mod $p(X)$. Then

$$p^\sigma(\xi) = p^\sigma(X^\sigma) = (p(X))^\sigma = 0.$$

Hence ξ is a root of p^σ, and as such is algebraic over σk. We have now found an extension of σk, namely $\sigma k(\xi)$ in which p^σ has a root.

With a minor set-theoretic argument, we shall have:

Proposition 2.3. *Let k be a field and f a polynomial in $k[X]$ of degree $\geqq 1$. Then there exists an extension E of k in which f has a root.*

Proof. We may assume that $f = p$ is irreducible. We have shown that there exists a field F and an embedding

$$\sigma: k \to F$$

such that p^σ has a root ξ in F. Let S be a set whose cardinality is the same as that of $F - \sigma k$ (= the complement of σk in F) and which is disjoint from k. Let $E = k \cup S$. We can extend $\sigma: k \to F$ to a bijection of E on F. We now define a field structure on E. If $x, y \in E$ we define

$$xy = \sigma^{-1}(\sigma(x)\sigma(y)),$$

$$x + y = \sigma^{-1}(\sigma(x) + \sigma(y)).$$

Restricted to k, our addition and multiplication coincide with the given addition and multiplication of our original field k, and it is clear that k is a subfield of E. We let $\alpha = \sigma^{-1}(\xi)$. Then it is also clear that $p(\alpha) = 0$, as desired.

Corollary 2.4. *Let k be a field and let $f_1, \ldots, f_n$ be polynomials in $k[X]$ of degrees $\geqq 1$. Then there exists an extension E of k in which each f_i has a root, $i = 1, \ldots, n$.*

Proof. Let E_1 be an extension in which f_1 has a root. We may view f_2 as a polynomial over E_1. Let E_2 be an extension of E_1 in which f_2 has a root. Proceeding inductively, our corollary follows at once.

We define a field L to be **algebraically closed** if every polynomial in $L[X]$ of degree $\geqq 1$ has a root in L.

Theorem 2.5. *Let k be a field. Then there exists an algebraically closed field containing k as a subfield.*

Proof. We first construct an extension E_1 of k in which every polynomial in $k[X]$ of degree $\geqq 1$ has a root. One can proceed as follows (Artin). To each polynomial f in $k[X]$ of degree $\geqq 1$ we associate a letter X_f and we let S be the set of all such letters X_f (so that S is in bijection with the set of polynomials in $k[X]$ of degree $\geqq 1$). We form the polynomial ring $k[S]$, and contend that the ideal generated by all the polynomials $f(X_f)$ in $k[S]$ is not the unit ideal. If it is, then there is a finite combination of elements in our ideal which is equal to 1:

$$g_1 f_1(X_{f_1}) + \cdots + g_n f_n(X_{f_n}) = 1$$

with $g_i \in k[S]$. For simplicity, write X_i instead of X_{f_i}. The polynomials g_i will involve actually only a finite number of variables, say $X_1, \ldots, X_N$ (with $N \geqq n$). Our relation then reads

$$\sum_{i=1}^{n} g_i(X_1, \ldots, X_N) f_i(X_i) = 1.$$

Let F be a finite extension in which each polynomial $f_1, \ldots, f_n$ has a root, say α_i is a root of f_i in F, for $i = 1, \ldots, n$. Let $\alpha_i = 0$ for $i > n$. Substitute α_i for X_i in our relation. We get $0 = 1$, contradiction.

Let $\mathfrak{m}$ be a maximal ideal containing the ideal generated by all polynomials $f(X_f)$ in $k[S]$. Then $k[S]/\mathfrak{m}$ is a field, and we have a canonical map

$$\sigma: k[S] \to k[S]/\mathfrak{m}.$$

For any polynomial $f \in k[X]$ of degree $\geqq 1$, the polynomial f^σ has a root in $k[S]/\mathfrak{m}$, which is an extension of σk. Using the same type of set-theoretic argument as in Proposition 2.3, we conclude that there exists an extension E_1 of k in which every polynomial $f \in k[X]$ of degree $\geqq 1$ has a root in E_1.

Inductively, we can form a sequence of fields

$$E_1 \subset E_2 \subset E_3 \subset \cdots \subset E_n \cdots$$

such that every polynomial in $E_n[X]$ of degree $\geqq 1$ has a root in E_{n+1}. Let E be the union of all fields E_n, $n = 1, 2, \ldots$. Then E is naturally a field, for if $x, y \in E$ then there exists some n such that $x, y \in E_n$, and we can take the product or sum xy or $x + y$ in E_n. This is obviously independent of the choice of n such that $x, y \in E_n$, and defines a field structure on E. Every polynomial in $E[X]$ has its coefficients in some subfield E_n, hence a root in E_{n+1}, hence a root in E, as desired.

Corollary 2.6. *Let k be a field. There exists an extension k^a which is algebraic over k and algebraically closed.*

Proof. Let E be an extension of k which is algebraically closed and let k^a be the union of all subextensions of E, which are algebraic over k. Then k^a is algebraic over k. If $\alpha \in E$ and α is algebraic over k^a then α is algebraic over k by Proposition 1.7. If f is a polynomial of degree $\geqq 1$ in $k^a[X]$, then f has a root α in E, and α is algebraic over k^a. Hence α is in k^a and k^a is algebraically closed.

We observe that if L is an algebraically closed field, and $f \in L[X]$ has degree $\geqq 1$, then there exists $c \in L$ and $\alpha_1, \ldots, \alpha_n \in L$ such that

$$f(X) = c(X - \alpha_1) \cdots (X - \alpha_n).$$

Indeed, f has a root α_1 in L, so there exists $g(X) \in L[X]$ such that

$$f(X) = (X - \alpha_1)g(X).$$

If $\deg g \geqq 1$, we can repeat this argument inductively, and express f as a

product of terms $(X - \alpha_i)$ $(i = 1, \dots, n)$ and an element $c \in L$. Note that c is the leading coefficient of f, i.e.

$$f(X) = cX^n + \text{terms of lower degree.}$$

Hence if the coefficients of f lie in a subfield k of L, then $c \in k$.

Let k be a field and $\sigma: k \to L$ an embedding of k into an algebraically closed field L. We are interested in analyzing the extensions of σ to algebraic extensions E of k. We begin by considering the special case when E is generated by one element.

Let $E = k(\alpha)$ where α is algebraic over k. Let

$$p(X) = \operatorname{Irr}(\alpha, k, X).$$

Let β be a root of p^σ in L. Given an element of $k(\alpha) = k[\alpha]$, we can write it in the form $f(\alpha)$ with some polynomial $f(X) \in k[X]$. We define an extension of σ by mapping

$$f(\alpha) \mapsto f^\sigma(\beta).$$

This is in fact well defined, i.e. independent of the choice of polynomial $f(X)$ used to express our element in $k[\alpha]$. Indeed, if $g(X)$ is in $k[X]$ and such that $g(\alpha) = f(\alpha)$, then $(g - f)(\alpha) = 0$, whence $p(X)$ divides $g(X) - f(X)$. Hence $p^\sigma(X)$ divides $g^\sigma(X) - f^\sigma(X)$, and thus $g^\sigma(\beta) = f^\sigma(\beta)$. It is now clear that our map is a homomorphism inducing σ on k, and that it is an extension of σ to $k(\alpha)$. Hence we get:

Proposition 2.7. *The number of possible extensions of σ to $k(\alpha)$ is $\leqq$ the number of roots of p, and is equal to the number of distinct roots of p.*

This is an important fact, which we shall analyze more closely later. For the moment, we are interested in extensions of σ to arbitrary algebraic extensions of k. We get them by using Zorn's lemma.

Theorem 2.8. *Let k be a field, E an algebraic extension of k, and $\sigma: k \to L$ an embedding of k into an algebraically closed field L. Then there exists an extension of σ to an embedding of E in L. If E is algebraically closed and L is algebraic over σk, then any such extension of σ is an isomorphism of E onto L.*

Proof. Let S be the set of all pairs (F, τ) where F is a subfield of E containing k, and τ is an extension of σ to an embedding of F in L. If (F, τ) and (F', τ') are such pairs, we write $(F, \tau) \leqq (F', \tau')$ if $F \subset F'$ and $\tau' | F = \tau$. Note that S is not empty [it contains (k, σ)], and is inductively ordered: If $\{(F_i, \tau_i)\}$ is a totally ordered subset, we let $F = \bigcup F_i$ and define τ on F to be equal to τ_i on each F_i. Then (F, τ) is an upper bound for the totally ordered subset. Using Zorn's lemma, let (K, λ) be a maximal element in S. Then λ is an extension of σ, and we contend that $K = E$. Otherwise, there exists $\alpha \in E$,

$\alpha \notin K$. By what we saw above, our embedding λ has an extension to $K(\alpha)$, thereby contradicting the maximality of (K, λ). This proves that there exists an extension of σ to E. We denote this extension again by σ.

If E is algebraically closed, and L is algebraic over σk, then σE is algebraically closed and L is algebraic over σE, hence $L = \sigma E$.

As a corollary, we have a certain uniqueness for an "algebraic closure" of a field k.

Corollary 2.9. *Let k be a field and let E, E' be algebraic extensions of k. Assume that E, E' are algebraically closed. Then there exists an isomorphism*

$$\tau: E \to E'$$

of E onto E' inducing the identity on k.

Proof. Extend the identity mapping on k to an embedding of E into E' and apply the theorem.

We see that an algebraically closed and algebraic extension of k is determined up to an isomorphism. Such an extension will be called an **algebraic closure** of k, and we frequently denote it by k^a. In fact, unless otherwise specified, we use the symbol k^a only to denote algebraic closure.

It is now worth while to recall the general situation of isomorphisms and automorphisms in general categories.

Let $\mathfrak{A}$ be a category, and A, B objects in $\mathfrak{A}$. We denote by $\mathrm{Iso}(A, B)$ the set of isomorphisms of A on B. Suppose there exists at least one such isomorphism $\sigma: A \to B$, with inverse $\sigma^{-1}: B \to A$. If φ is an automorphism of A, then $\sigma \circ \varphi: A \to B$ is again an isomorphism. If ψ is an automorphism of B, then $\psi \circ \sigma: A \to B$ is again an isomorphism. Furthermore, the groups of automorphisms $\mathrm{Aut}(A)$ and $\mathrm{Aut}(B)$ are isomorphic, under the mappings

$$\varphi \mapsto \sigma \circ \varphi \circ \sigma^{-1},$$

$$\sigma^{-1} \circ \psi \circ \sigma \leftarrow\!\shortmid \psi,$$

which are inverse to each other. The isomorphism $\sigma \circ \varphi \circ \sigma^{-1}$ is the one which makes the following diagram commutative:

$$\begin{array}{ccc} A & \xrightarrow{\sigma} & B \\ {\scriptstyle \varphi}\downarrow & & \downarrow{\scriptstyle \sigma\circ\varphi\circ\sigma^{-1}} \\ A & \xrightarrow[\sigma]{} & B \end{array}$$

We have a similar diagram for $\sigma^{-1} \circ \psi \circ \sigma$.

Let $\tau: A \to B$ be another isomorphism. Then $\tau^{-1} \circ \sigma$ is an automorphism of A, and $\tau \circ \sigma^{-1}$ is an automorphism of B. Thus two isomorphisms differ by an automorphism (of A or B). We see that the group $\mathrm{Aut}(B)$ operates on the

set $\text{Iso}(A, B)$ on the left, and $\text{Aut}(A)$ operates on the set $\text{Iso}(A, B)$ on the right.

We also see that $\text{Aut}(A)$ is determined up to a mapping analogous to a conjugation. This is quite different from the type of uniqueness given by universal objects in a category. Such objects have only the identity automorphism, and hence are determined up to a unique isomorphism.

This is not the case with the algebraic closure of a field, which usually has a large amount of automorphisms. Most of this chapter and the next is devoted to the study of such automorphisms.

Examples. It will be proved later in this book that the complex numbers are algebraically closed. Complex conjugation is an automorphism of $\mathbf{C}$. There are many more automorphisms, but the other automorphisms $\neq$ id. are not continuous. We shall discuss other possible automorphisms in the chapter on transcendental extensions. The subfield of $\mathbf{C}$ consisting of all numbers which are algebraic over $\mathbf{Q}$ is an algebraic closure $\mathbf{Q}^{\mathrm{a}}$ of $\mathbf{Q}$. It is easy to see that $\mathbf{Q}^{\mathrm{a}}$ is denumerable. In fact, prove the following as an exercise:

If k is a field which is not finite, then any algebraic extension of k has the same cardinality as k.

If k is denumerable, one can first enumerate all polynomials in k, then enumerate finite extensions by their degree, and finally enumerate the cardinality of an arbitrary algebraic extension. We leave the counting details as exercises.

In particular, $\mathbf{Q}^{\mathrm{a}} \neq \mathbf{C}$. If $\mathbf{R}$ is the field of real numbers, then $\mathbf{R}^{\mathrm{a}} = \mathbf{C}$.

If k is a finite field, then algebraic closure k^{a} of k is denumerable. We shall in fact describe in great detail the nature of algebraic extensions of finite fields later in this chapter.

Not all interesting fields are subfields of the complex numbers. For instance, one wants to investigate the algebraic extensions of a field $\mathbf{C}(X)$ where X is a variable over $\mathbf{C}$. The study of these extensions amounts to the study of ramified coverings of the sphere (viewed as a Riemann surface), and in fact one has precise information concerning the nature of such extensions, because one knows the fundamental group of the sphere from which a finite number of points has been deleted. We shall mention this example again later when we discuss Galois groups.

§3. SPLITTING FIELDS AND NORMAL EXTENSIONS

Let k be a field and let f be a polynomial in $k[X]$ of degree $\geqq 1$. By a **splitting field** K of f we shall mean an extension K of k such that f splits into linear factors in K, i.e.

$$f(X) = c(X - \alpha_1) \cdots (X - \alpha_n)$$

with $\alpha_i \in K$, $i = 1, \ldots, n$, and such that $K = k(\alpha_1, \ldots, \alpha_n)$ is generated by all the roots of f.

Theorem 3.1. *Let K be a splitting field of the polynomial $f(X) \in k[X]$. If E is another splitting field of f, then there exists an isomorphism $\sigma: E \to K$ inducing the identity on k. If $k \subset K \subset k^a$, where k^a is an algebraic closure of k, then any embedding of E in k^a inducing the identity on k must be an isomorphism of E onto K.*

Proof. Let K^a be an algebraic closure of K. Then K^a is algebraic over k, hence is an algebraic closure of k. By Theorem 2.8 there exists an embedding

$$\sigma: E \to K^a$$

inducing the identity on k. We have a factorization

$$f(X) = c(X - \beta_1) \cdots (X - \beta_n)$$

with $\beta_i \in E$, $i = 1, \ldots, n$. The leading coefficient c lies in k. We obtain

$$f(X) = f^\sigma(X) = c(X - \sigma\beta_1) \cdots (X - \sigma\beta_n).$$

We have unique factorization in $K^a[X]$. Since f has a factorization

$$f(X) = c(X - \alpha_1) \cdots (X - \alpha_n)$$

in $K[X]$, it follows that $(\sigma\beta_1, \ldots, \sigma\beta_n)$ differs from $(\alpha_1, \ldots, \alpha_n)$ by a permutation. From this we conclude that $\sigma\beta_i \in K$ for $i = 1, \ldots, n$ and hence that $\sigma E \subset K$. But $K = k(\alpha_1, \ldots, \alpha_n) = k(\sigma\beta_1, \ldots, \sigma\beta_n)$, and hence $\sigma E = K$, because

$$E = k(\beta_1, \ldots, \beta_n).$$

This proves our theorem.

We note that a polynomial $f(X) \in k[X]$ always has a splitting field, namely the field generated by its roots in a given algebraic closure k^a of k.

Let I be a set of indices and let $\{f_i\}_{i \in I}$ be a family of polynomials in $k[X]$, of degrees $\geqq 1$. By a **splitting field** for this family we shall mean an extension K of k such that every f_i splits in linear factors in $K[X]$, and K is generated by all the roots of all the polynomials f_i, $i \in I$. In most applications we deal with a finite indexing set I, but it is becoming increasingly important to consider infinite algebraic extensions, and so we shall deal with them fairly systematically. One should also observe that the proofs we shall give for various statements would not be simpler if we restricted ourselves to the finite case.

Let k^a be an algebraic closure of k, and let K_i be a splitting field of f_i in k^a. Then the compositum of the K_i is a splitting field for our family,

since the two conditions defining a splitting field are immediately satisfied. Furthermore Theorem 3.1 extends at once to the infinite case:

Corollary 3.2. *Let K be a splitting field for the family $\{f_i\}_{i \in I}$ and let E be another splitting field. Any embedding of E into K^a inducing the identity on k gives an isomorphism of E onto K.*

Proof. Let the notation be as above. Note that E contains a unique splitting field E_i of f_i and K contains a unique splitting field K_i of f_i. Any embedding σ of E into K^a must map E_i onto K_i by Theorem 3.1, and hence maps E into K. Since K is the compositum of the fields K_i, our map σ must send E onto K and hence induces an isomorphism of E onto K.

Remark. If I is finite, and our polynomials are $f_1, \ldots, f_n$, then a splitting field for them is a splitting field for the single polynomial $f(X) = f_1(X) \cdots f_n(X)$ obtained by taking the product. However, even when dealing with finite extensions only, it is convenient to deal simultaneously with sets of polynomials rather than a single one.

Theorem 3.3. *Let K be an algebraic extension of k, contained in an algebraic closure k^a of k. Then the following conditions are equivalent:*

NOR 1. *Every embedding of K in k^a over k induces an automorphism of K.*

NOR 2. *K is the splitting field of a family of polynomials in $k[X]$.*

NOR 3. *Every irreducible polynomial of $k[X]$ which has a root in K splits into linear factors in K.*

Proof. Assume **NOR 1**. Let α be an element of K and let $p_\alpha(X)$ be its irreducible polynomial over k. Let β be a root of p_α in k^a. There exists an isomorphism of $k(\alpha)$ on $k(\beta)$ over k, mapping α on β. Extend this isomorphism to an embedding of K in k^a. This extension is an automorphism σ of K by hypothesis, hence $\sigma\alpha = \beta$ lies in K. Hence every root of p_α lies in K, and p_α splits in linear factors in $K[X]$. Hence K is the splitting field of the family $\{p_\alpha\}_{\alpha \in K}$ as α ranges over all elements of K, and **NOR 2** is satisfied.

Conversely, assume **NOR 2**, and let $\{f_i\}_{i \in I}$ be the family of polynomials of which K is the splitting field. If α is a root of some f_i in K, then for any embedding σ of K in k^a over k we know that $\sigma\alpha$ is a root of f_i. Since K is generated by the roots of all the polynomials f_i, it follows that σ maps K into itself. We now apply Lemma 2.1 to conclude that σ is an automorphism.

Our proof that **NOR 1** implies **NOR 2** also shows that **NOR 3** is satisfied. Conversely, assume **NOR 3**. Let σ be an embedding of K in k^a over k. Let $\alpha \in K$ and let $p(X)$ be its irreducible polynomial over k. If σ is an embedding of K in k^a over k then σ maps α on a root β of $p(X)$, and by hypothesis β lies in K. Hence $\sigma\alpha$ lies in K, and σ maps K into itself. By Lemma 2.1, it follows that σ is an automorphism.

An extension K of k satisfying the hypotheses **NOR 1, NOR 2, NOR 3** will be said to be **normal**. It is not true that the class of normal extensions is distinguished. For instance, it is easily shown that an extension of degree 2 is normal, but the extension $\mathbf{Q}(\sqrt[4]{2})$ of the rational numbers is not normal (the complex roots of $X^4 - 2$ are not in it), and yet this extension is obtained by successive extensions of degree 2, namely

$$E = \mathbf{Q}(\sqrt[4]{2}) \supset F \supset \mathbf{Q},$$

where

$$F = \mathbf{Q}(\alpha), \qquad \alpha = \sqrt{2} \quad \text{and} \quad E = F(\sqrt{\alpha}).$$

Thus a tower of normal extensions is not necessarily normal. However, we still have some of the properties:

Theorem 3.4. *Normal extensions remain normal under lifting. If $K \supset E \supset k$ and K is normal over k, then K is normal over E. If K_1, K_2 are normal over k and are contained in some field L, then K_1K_2 is normal over k, and so is $K_1 \cap K_2$.*

Proof. For our first assertion, let K be normal over k, let F be any extension of k, and assume K, F are contained in some bigger field. Let σ be an embedding of KF over F (in F^a). Then σ induces the identity on F, hence on k, and by hypothesis its restriction to K maps K into itself. We get $(KF)^\sigma = K^\sigma F^\sigma = KF$ whence KF is normal over F.

Assume that $K \supset E \supset k$ and that K is normal over k. Let σ be an embedding of K over E. Then σ is also an embedding of K over k, and our assertion follows by definition.

Finally, if K_1, K_2 are normal over k, then for any embedding σ of K_1K_2 over k we have

$$\sigma(K_1K_2) = \sigma(K_1)\sigma(K_2)$$

and our assertion again follows from the hypothesis. The assertion concerning the intersection is true because

$$\sigma(K_1 \cap K_2) = \sigma(K_1) \cap \sigma(K_2).$$

We observe that if K is a finitely generated normal extension of k, say

$$K = k(\alpha_1, \ldots, \alpha_n),$$

and $p_1, \ldots, p_n$ are the respective irreducible polynomials of $\alpha_1, \ldots, \alpha_n$ over k then K is already the splitting field of the finite family $p_1, \ldots, p_n$. We shall investigate later when K is the splitting field of a single irreducible polynomial.

§4. SEPARABLE EXTENSIONS

Let E be an algebraic extension of a field F and let

$$\sigma: F \to L$$

be an embedding of F in an algebraically closed field L. We investigate more closely extensions of σ to E. Any such extension of σ maps E on a subfield of L which is algebraic over σF. Hence for our purposes, we shall assume that L is algebraic over σF, hence is equal to an algebraic closure of σF.

Let S_σ be the set of extensions of σ to an embedding of E in L.

Let L' be another algebraically closed field, and let $\tau: F \to L'$ be an embedding. We assume as before that L' is an algebraic closure of τF. By Theorem 2.8, there exists an isomorphism $\lambda: L \to L'$ extending the map $\tau \circ \sigma^{-1}$ applied to the field σF. This is illustrated in the following diagram:

$$\begin{array}{ccccc} L' & \xleftarrow{\quad\lambda\quad} & & & L \\ | & \longleftarrow & E & \xrightarrow{\sigma^*} & | \\ | & & | & & | \\ \tau F & \xleftarrow[\tau]{} & F & \xrightarrow[\sigma]{} & \sigma F \end{array}$$

We let S_τ be the set of embeddings of E in L' extending τ.

If $\sigma^* \in S_\sigma$ is an extension of σ to an embedding of E in L, then $\lambda \circ \sigma^*$ is an extension of τ to an embedding of E into L', because for the restriction to F we have

$$\lambda \circ \sigma^* = \tau \circ \sigma^{-1} \circ \sigma = \tau.$$

Thus λ induces a mapping from S_σ into S_τ. It is clear that the inverse mapping is induced by λ^{-1}, and hence that S_σ, S_τ are in bijection under the mapping

$$\sigma^* \mapsto \lambda \circ \sigma^*.$$

In particular, the cardinality of S_σ, S_τ is the same. Thus this cardinality depends only on the extension E/F, and will be denoted by

$$[E : F]_s.$$

We shall call it the **separable degree** of E over F. It is mostly interesting when E/F is finite.

Theorem 4.1. *Let $E \supset F \supset k$ be a tower. Then*

$$[E : k]_s = [E : F]_s [F : k]_s.$$

Furthermore, if E is finite over k, then $[E : k]_s$ is finite and

$$[E:k]_s \leqq [E:k].$$

The separable degree is at most equal to the degree.

Proof. Let $\sigma: k \to L$ be an embedding of k in an algebraically closed field L. Let $\{\sigma_i\}_{i \in I}$ be the family of distinct extensions of σ to F, and for each i, let $\{\tau_{ij}\}$ be the family of distinct extensions of σ_i to E. By what we saw before, each σ_i has precisely $[E:F]_s$ extensions to embeddings of E in L. The set of embeddings $\{\tau_{ij}\}$ contains precisely

$$[E:F]_s[F:k]_s$$

elements. Any embedding of E into L over σ must be one of the τ_{ij}, and thus we see that the first formula holds, i.e. we have multiplicativity in towers.

As to the second, let us assume that E/k is finite. Then we can obtain E as a tower of extensions, each step being generated by one element:

$$k \subset k(\alpha_1) \subset k(\alpha_1, \alpha_2) \subset \cdots \subset k(\alpha_1, \ldots, \alpha_r) = E.$$

If we define inductively $F_{v+1} = F_v(\alpha_{v+1})$ then by Proposition 2.7,

$$[F_v(\alpha_{v+1}) : F_v]_s \leqq [F_v(\alpha_{v+1}) : F_v].$$

Thus our inequality is true in each step of the tower. By multiplicativity, it follows that the inequality is true for the extension E/k, as was to be shown.

Corollary 4.2. *Let E be finite over k, and $E \supset F \supset k$. The equality*

$$[E:k]_s = [E:k]$$

holds if and only if the corresponding equality holds in each step of the tower, i.e. for E/F and F/k.

Proof. Clear.

It will be shown later (and it is not difficult to show) that $[E:k]_s$ divides the degree $[E:k]$ when E is finite over k. We define $[E:k]_i$ to be the quotient, so that

$$[E:k]_s[E:k]_i = [E:k].$$

It then follows from the multiplicativity of the separable degree and of the degree in towers that the symbol $[E:k]_i$ is also multiplicative in towers. We shall deal with it at greater length in §6.

Let E be a finite extension of k. We shall say that E is **separable** over k if $[E:k]_s = [E:k]$.

An element α algebraic over k is said to be **separable** over k if $k(\alpha)$ is separable over k. We see that this condition is equivalent to saying that the irreducible polynomial $\mathrm{Irr}(\alpha, k, X)$ has no multiple roots.

A polynomial $f(X) \in k[X]$ is called **separable** if it has no multiple roots.

If α is a root of a separable polynomial $g(X) \in k[X]$ then the irreducible polynomial of α over k divides g and hence α is separable over k.

We note that if $k \subset F \subset K$ and $\alpha \in K$ is separable over k, then α is separable over F. Indeed, if f is a separable polynomial in $k[X]$ such that $f(\alpha) = 0$, then f also has coefficients in F, and thus α is separable over F. (We may say that a separable element remains separable under lifting.)

Theorem 4.3. *Let E be a finite extension of k. Then E is separable over k if and only if each element of E is separable over k.*

Proof. Assume E is separable over k and let $\alpha \in E$. We consider the tower

$$k \subset k(\alpha) \subset E.$$

By Corollary 4.2, we must have $[k(\alpha):k] = [k(\alpha):k]_s$ whence α is separable over k. Conversely, assume that each element of E is separable over k. We can write $E = k(\alpha_1, \ldots, \alpha_n)$ where each α_i is separable over k. We consider the tower

$$k \subset k(\alpha_1) \subset k(\alpha_1, \alpha_2) \subset \cdots \subset k(\alpha_1, \ldots, \alpha_n).$$

Since each α_i is separable over k, each α_i is separable over $k(\alpha_1, \ldots, \alpha_{i-1})$ for $i \geqq 2$. Hence by the tower theorem, it follows that E is separable over k.

We observe that our last argument shows: If E is generated by a finite number of elements, each of which is separable over k, then E is separable over k.

Let E be an arbitrary algebraic extension of k. We define E to be **separable** over k if every finitely generated subextension is separable over k, i.e., if every extension $k(\alpha_1, \ldots, \alpha_n)$ with $\alpha_1, \ldots, \alpha_n \in E$ is separable over k.

Theorem 4.4. *Let E be an algebraic extension of k, generated by a family of elements $\{\alpha_i\}_{i \in I}$. If each α_i is separable over k then E is separable over k.*

Proof. Every element of E lies in some finitely generated subfield

$$k(\alpha_{i_1}, \ldots, a_{i_n}),$$

and as we remarked above, each such subfield is separable over k. Hence every element of E is separable over k by Theorem 4.3, and this concludes the proof.

Theorem 4.5. *Separable extensions form a distinguished class of extensions.*

Proof. Assume that E is separable over k and let $E \supset F \supset k$. Every element of E is separable over F, and every element of F is an element of E, so separable over k. Hence each step in the tower is separable. Conversely, assume that $E \supset F \supset k$ is some extension such that E/F is separable and F/k is separable. If E is finite over k, then we can use Corollary 4.2. Namely, we have an equality of the separable degree and the degree in each step of the tower, whence an equality for E over k by multiplicativity.

If E is infinite, let $\alpha \in E$. Then α is a root of a separable polynomial $f(X)$ with coefficients in F. Let these coefficients be $a_n, \ldots, a_0$. Let $F_0 = k(a_n, \ldots, a_0)$. Then F_0 is separable over k, and α is separable over F_0. We now deal with the finite tower

$$k \subset F_0 \subset F_0(\alpha)$$

and we therefore conclude that $F_0(\alpha)$ is separable over k, hence that α is separable over k. This proves condition (1) in the definition of "distinguished."

Let E be separable over k. Let F be any extension of k, and assume that E, F are both subfields of some field. Every element of E is separable over k, whence separable over F. Since EF is generated over F by all the elements of E, it follows that EF is separable over F, by Theorem 4.4. This proves condition (2) in the definition of "distinguished," and concludes the proof of our theorem.

Let E be a finite extension of k. The intersection of all normal extensions K of k (in an algebraic closure E^a) containing E is a normal extension of k which contains E, and is obviously the smallest normal extension of k containing E. If $\sigma_1, \ldots, \sigma_n$ are the distinct embeddings of E in E^a, then the extension

$$K = (\sigma_1 E)(\sigma_2 E) \cdots (\sigma_n E),$$

which is the compositum of all these embeddings, is a normal extension of k, because for any embedding of it, say τ, we can apply τ to each extension $\sigma_i E$. Then $(\tau\sigma_1, \ldots, \tau\sigma_n)$ is a permutation of $(\sigma_1, \ldots, \sigma_n)$ and thus τ maps K into itself. Any normal extension of k containing E must contain $\sigma_i E$ for each i, and thus *the smallest normal extension of k containing E is precisely equal to the compositum*

$$(\sigma_1 E) \cdots (\sigma_n E).$$

If E is separable over k, then from Theorem 4.5 and induction we conclude that the smallest normal extension of k containing E is also separable over k.

Similar results hold for an infinite algebraic extension E of k, taking an infinite compositum.

In light of Theorem 4.5, the compositum of all separable extensions of a field k in a given algebraic closure k^a is a separable extension, which will be denoted by k^s or k^{sep}, and will be called the **separable closure** of k. As a matter of terminology, if E is an algebraic extension of k, and σ any embedding of E in k^a over k, then we call σE a **conjugate** of E in k^a. *We can say that the smallest normal extension of k containing E is the compositum of all the conjugates of E in E^a.*

Let α be algebraic over k. If $\sigma_1, \ldots, \sigma_r$ are the distinct embeddings of $k(\alpha)$ into k^a over k, then we call $\sigma_1\alpha, \ldots, \sigma_r\alpha$ the **conjugates** of α in k^a. These elements are simply the distinct roots of the irreducible polynomial of α over k. The smallest normal extension of k containing one of these conjugates is simply $k(\sigma_1\alpha, \ldots, \sigma_r\alpha)$.

Theorem 4.6. (Primitive Element Theorem). *Let E be a finite extension of a field k. There exists an element $\alpha \in E$ such that $E = k(\alpha)$ if and only if there exists only a finite number of fields F such that $k \subset F \subset E$. If E is separable over k, then there exists such an element α.*

Proof. If k is finite, then we know that the multiplicative group of E is generated by one element, which will therefore also generate E over k. We assume that k is infinite.

Assume that there is only a finite number of fields, intermediate between k and E. Let $\alpha, \beta \in E$. As c ranges over elements of k, we can only have a finite number of fields of type $k(\alpha + c\beta)$. Hence there exist elements c_1, $c_2 \in k$ with $c_1 \neq c_2$ such that

$$k(\alpha + c_1\beta) = k(\alpha + c_2\beta).$$

Note that $\alpha + c_1\beta$ and $\alpha + c_2\beta$ are in the same field, whence so is $(c_1 - c_2)\beta$, and hence so is β. Thus α is also in that field, and we see that $k(\alpha, \beta)$ can be generated by one element.

Proceeding inductively, if $E = k(\alpha_1, \ldots, \alpha_n)$ then there will exist elements $c_2, \ldots, c_n \in k$ such that

$$E = k(\xi)$$

where $\xi = \alpha_1 + c_2\alpha_2 + \cdots + c_n\alpha_n$. This proves half of our theorem.

Conversely, assume that $E = k(\alpha)$ for some α, and let $f(X) = \operatorname{Irr}(\alpha, k, X)$. Let $k \subset F \subset E$. Let $g_F(X) = \operatorname{Irr}(\alpha, F, X)$. Then g_F divides f. We have unique factorization in $E[X]$, and any polynomial in $E[X]$ which has leading coefficient 1 and divides $f(X)$ is equal to a product of factors $(X - \alpha_i)$ where $\alpha_1, \ldots, \alpha_n$ are the roots of f in a fixed algebraic closure. Hence there is only a finite number of such polynomials. Thus we get a mapping

$$F \mapsto g_F$$

from the set of intermediate fields into a finite set of polynomials. Let F_0 be

the subfield of F generated over k by the coefficients of $g_F(X)$. Then g_F has coefficients in F_0 and is irreducible over F_0 since it is irreducible over F. Hence the degree of α over F_0 is the same as the degree of α over F. Hence $F = F_0$. Thus our field F is uniquely determined by its associated polynomials g_F, and our mapping is therefore injective. This proves the first assertion of the theorem.

As to the statement concerning separable extensions, using induction, we may assume without loss of generality that $E = k(\alpha, \beta)$ where α, β are separable over k. Let $\sigma_1, \dots, \sigma_n$ be the distinct embeddings of $k(\alpha, \beta)$ in k^a over k. Let

$$P(X) = \prod_{i \neq j} (\sigma_i\alpha + X\sigma_i\beta - \sigma_j\alpha - X\sigma_j\beta).$$

Then $P(X)$ is not the zero polynomial, and hence there exists $c \in k$ such that $P(c) \neq 0$. Then the elements $\sigma_i(\alpha + c\beta)$ $(i = 1, \dots, n)$ are distinct, whence $k(\alpha + c\beta)$ has degree at least n over k. But $n = [k(\alpha, \beta) : k]$, and hence

$$k(\alpha, \beta) = k(\alpha + c\beta),$$

as desired.

If $E = k(\alpha)$, then we say that α is a **primitive element** of E (over k).

§5. FINITE FIELDS

We have developed enough general theorems to describe the structure of finite fields. This is interesting for its own sake, and also gives us examples for the general theory.

Let F be a finite field with q elements. As we have noted previously, we have a homomorphism

$$\mathbf{Z} \to F$$

sending 1 on 1, whose kernel cannot be 0, and hence is a principal ideal generated by a prime number p since $\mathbf{Z}/p\mathbf{Z}$ is embedded in F and F has no divisors of zero. Thus F has characteristic p, and contains a field isomorphic to $\mathbf{Z}/p\mathbf{Z}$.

We remark that $\mathbf{Z}/p\mathbf{Z}$ has no automorphisms other than the identity. Indeed, any automorphism must map 1 on 1, hence leaves every element fixed because 1 generates $\mathbf{Z}/p\mathbf{Z}$ additively. We identify $\mathbf{Z}/p\mathbf{Z}$ with its image in F. Then F is a vector space over $\mathbf{Z}/p\mathbf{Z}$, and this vector space must be

finite since F is finite. Let its degree be n. Let $\omega_1, \dots, \omega_n$ be a basis for F over $\mathbf{Z}/p\mathbf{Z}$. Every element of F has a unique expression of the form

$$a_1\omega_1 + \cdots + a_n\omega_n$$

with $a_i \in \mathbf{Z}/p\mathbf{Z}$. Hence $q = p^n$.

The multiplicative group F^* of F has order $q - 1$. Every $\alpha \in F^*$ satisfies the equation $X^{q-1} = 1$. Hence every element of F satisfies the equation

$$f(X) = X^q - X = 0.$$

This implies that the polynomial $f(X)$ has q distinct roots in F, namely all elements of F. Hence f splits into factors of degree 1 in F, namely

$$X^q - X = \prod_{\alpha \in F} (X - \alpha).$$

In particular, F is a splitting field for f. But a splitting field is uniquely determined up to an isomorphism. Hence if a finite field of order p^n exists, it is uniquely determined, up to an isomorphism, as the splitting field of $X^{p^n} - X$ over $\mathbf{Z}/p\mathbf{Z}$.

As a matter of notation, we denote $\mathbf{Z}/p\mathbf{Z}$ by $\mathbf{F}_p$. Let n be an integer $\geqq 1$ and consider the splitting field of

$$X^{p^n} - X = f(X)$$

in an algebraic closure $\mathbf{F}_p^{\mathrm{a}}$. We contend that this splitting field is the set of roots of $f(X)$ in $\mathbf{F}_p^{\mathrm{a}}$. Indeed, let α, β be roots. Then

$$(\alpha + \beta)^{p^n} - (\alpha + \beta) = \alpha^{p^n} + \beta^{p^n} - \alpha - \beta = 0,$$

whence $\alpha + \beta$ is a root. Also,

$$(\alpha\beta)^{p^n} - \alpha\beta = \alpha^{p^n}\beta^{p^n} - \alpha\beta = \alpha\beta - \alpha\beta = 0,$$

and $\alpha\beta$ is a root. Note that 0, 1 are roots of $f(X)$. If $\beta \neq 0$ then

$$(\beta^{-1})^{p^n} - \beta^{-1} = (\beta^{p^n})^{-1} - \beta^{-1} = 0$$

so that β^{-1} is a root. Finally,

$$(-\beta)^{p^n} - (-\beta) = (-1)^{p^n}\beta^{p^n} + \beta.$$

If p is odd, then $(-1)^{p^n} = -1$ and we see that $-\beta$ is a root. If p is even then $-1 = 1$ (in $\mathbf{Z}/2\mathbf{Z}$) and hence $-\beta = \beta$ is a root. This proves our contention.

The derivative of $f(X)$ is

$$f'(X) = p^n X^{p^n - 1} - 1 = -1.$$

Hence $f(X)$ has no multiple roots, and therefore has p^n distinct roots in $\mathbf{F}_p^{\mathrm{a}}$. Hence its splitting field has exactly p^n elements. We summarize our results:

Theorem 5.1. *For each prime p and each integer $n \geqq 1$ there exists a finite field of order p^n denoted by $\mathbf{F}_{p^n}$, uniquely determined as a subfield of an algebraic closure $\mathbf{F}_p^a$. It is the splitting field of the polynomial*

$$X^{p^n} - X,$$

and its elements are the roots of this polynomial. Every finite field is isomorphic to exactly one field $\mathbf{F}_{p^n}$.

We usually write $p^n = q$ and $\mathbf{F}_q$ instead of $\mathbf{F}_{p^n}$.

Corollary 5.2. *Let $\mathbf{F}_q$ be a finite field. Let n be an integer $\geqq 1$. In a given algebraic closure $\mathbf{F}_q^a$, there exists one and only one extension of $\mathbf{F}_q$ of degree n, and this extension is the field $\mathbf{F}_{q^n}$.*

Proof. Let $q = p^m$. Then $q^n = p^{mn}$. The splitting field of $X^{q^n} - X$ is precisely $F_{p^{mn}}$ and has degree mn over $\mathbf{Z}/p\mathbf{Z}$. Since $\mathbf{F}_q$ has degree m over $\mathbf{Z}/p\mathbf{Z}$, it follows that $\mathbf{F}_{q^n}$ has degree n over $\mathbf{F}_q$. Conversely, any extension of degree n over $\mathbf{F}_q$ has degree mn over $\mathbf{F}_p$ and hence must be $\mathbf{F}_{p^{mn}}$. This proves our corollary.

Theorem 5.3. *The multiplicative group of a finite field is cyclic.*

Proof. This has already been proved in Chapter IV, Theorem 1.9.

We shall determine all automorphisms of a finite field.

Let $q = p^n$ and let $\mathbf{F}_q$ be the finite field with q elements. We consider the **Frobenius mapping**

$$\varphi: \mathbf{F}_q \to \mathbf{F}_q$$

such that $\varphi(x) = x^p$. Then φ is a homomorphism, and its kernel is 0 since $\mathbf{F}_q$ is a field. Hence φ is injective. Since $\mathbf{F}_q$ is finite, it follows that φ is surjective, and hence that φ is an isomorphism. We note that it leaves $\mathbf{F}_p$ fixed.

Theorem 5.4. *The group of automorphisms of $\mathbf{F}_q$ is cyclic of degree n, generated by φ.*

Proof. Let G be the group generated by φ. We note that $\varphi^n = \text{id}$ because $\varphi^n(x) = x^{p^n} = x$ for all $x \in \mathbf{F}_q$. Hence n is an exponent for φ. Let d be the period of φ, so $d \geqq 1$. We have $\varphi^d(x) = x^{p^d}$ for all $x \in \mathbf{F}_q$. Hence each $x \in \mathbf{F}_q$ is a root of the equation

$$X^{p^d} - X = 0.$$

This equation has at most p^d roots. It follows that $d \geqq n$, whence $d = n$.

There remains to be proved that G is the group of all automorphisms of $\mathbf{F}_q$. Any automorphism of $\mathbf{F}_q$ must leave $\mathbf{F}_p$ fixed. Hence it is an auto-

morphism of $\mathbf{F}_q$ over $\mathbf{F}_p$. By Theorem 4.1, the number of such automorphisms is $\leqq n$. Hence $\mathbf{F}_q$ cannot have any other automorphisms except for those of G.

Theorem 5.5. *Let m, n be integers $\geqq 1$. Then in any algebraic closure of $\mathbf{F}_p$, the subfield $\mathbf{F}_{p^n}$ is contained in $\mathbf{F}_{p^m}$ if and only if n divides m. If that is the case, let $q = p^n$, and let $m = nd$. Then $\mathbf{F}_{p^m}$ is normal and separable over $\mathbf{F}_q$, and the group of automorphisms of $\mathbf{F}_{pm}$ over $\mathbf{F}_q$ is cyclic of order d, generated by φ^n.*

Proof. All the statements are trivial consequences of what has already been proved and will be left to the reader.

§6. INSEPARABLE EXTENSIONS

This section is of a fairly technical nature, and can be omitted without impairing the understanding of most of the rest of the book.

We begin with some remarks supplementing those of Proposition 2.7.

Let $f(X) = (X - \alpha)^m g(X)$ be a polynomial in $k[X]$, and assume $X - \alpha$ does not divide $g(X)$. We recall that m is called the multiplicity of α in f. We say that α is a **multiple** root of f if $m > 1$. Otherwise, we say that α is a **simple** root.

Proposition 6.1. *Let α be algebraic over k, $\alpha \in k^a$, and let*

$$f(X) = \mathrm{Irr}(\alpha, k, X).$$

If char $k = 0$, *then all roots of f have multiplicity* 1 *(f is separable). If*

$$\mathrm{char}\, k = p > 0,$$

then there exists an integer $\mu \geqq 0$ such that every root of f has multiplicity p^μ. We have

$$[k(\alpha) : k] = p^\mu [k(\alpha) : k]_s,$$

and α^{p^μ} is separable over k.

Proof. Let $\alpha_1, \ldots, \alpha_r$ be the distinct roots of f in k^a and let $\alpha = \alpha_1$. Let m be the multiplicity of α in f. Given $1 \leqq i \leqq r$, there exists an isomorphism

$$\sigma: k(\alpha) \to k(\alpha_i)$$

over k such that $\sigma\alpha = \alpha_i$. Extend σ to an automorphism of k^a and denote

this extension also by σ. Since f has coefficients in k we have $f^\sigma = f$. We note that

$$f(X) = \prod_{j=1}^{r} (X - \sigma\alpha_j)^{m_j}$$

if m_j is the multiplicity of α_j in f. By unique factorization, we conclude that $m_i = m_1$ and hence that all m_i are equal to the same integer m.

Consider the derivative $f'(X)$. If f and f' have a root in common, then α is a root of a polynomial of lower degree than $\deg f$. This is impossible unless $\deg f' = -\infty$, in other words, f' is identically 0. If the characteristic is 0, this cannot happen. Hence if f has multiple roots, we are in characteristic p, and $f(X) = g(X^p)$ for some polynomial $g(X) \in k[X]$. Therefore α^p is a root of a polynomial g whose degree is $< \deg f$. Proceeding inductively, we take the smallest integer $\mu \geqq 0$ such that α^{p^μ} is the root of a separable polynomial in $k[X]$, namely the polynomial h such that

$$f(X) = h(X^{p^\mu}).$$

Comparing the degree of f and g, we conclude that

$$[k(\alpha) : k(\alpha^p)] = p.$$

Inductively, we find

$$[k(\alpha) : k(\alpha^{p^\mu})] = p^\mu.$$

Since h has roots of multiplicity 1, we know that

$$[k(\alpha^{p^\mu}) : k]_s = [k(\alpha^{p^\mu}) : k],$$

and comparing the degree of f and the degree of h, we see that the number of distinct roots of f is equal to the number of distinct roots of h. Hence

$$[k(\alpha) : k]_s = [k(\alpha^{p^\mu}) : k]_s.$$

From this our formula for the degree follows by multiplicativity, and our proposition is proved. We note that the roots of h are

$$\alpha_1^{p^\mu}, \ldots, \alpha_r^{p^\mu}.$$

Corollary 6.2. *For any finite extension E of k, the separable degree $[E : k]_s$ divides the degree $[E : k]$. The quotient is* 1 *if the characteristic is* 0, *and a power of p if the characteristic is $p > 0$.*

Proof. We decompose E/k into a tower, each step being generated by one element, and apply Proposition 6.1, together with the multiplicativity of our indices in towers.

If E/K is finite, we call the quotient

$$\frac{[E:k]}{[E:k]_s}$$

the **inseparable degree** (or **degree of inseparability**), and denote it by $[E:k]_i$ as in §4. We have

$$[E:k]_s[E:k]_i = [E:k].$$

Corollary 6.3. *A finite extension is separable if and only if* $[E:k]_i = 1$.

Proof. By definition.

Corollary 6.4 *If* $E \supset F \supset k$ *are two finite extensions, then*

$$[E:k]_i = [E:F]_i[F:k]_i.$$

Proof. Immediate by Theorem 4.1.

We now assume throughout that k is a field of characteristic $p > 0$.

An element α algebraic over k is said to be **purely inseparable** over k if there exists an integer $n \geqq 0$ such that α^{p^n} lies in k.

Let E be an algebraic extension of k. We contend that the following conditions are equivalent:

P. Ins. 1. We have $[E:k]_s = 1$.

P. Ins. 2. Every element α of E is purely inseparable over k.

P. Ins. 3. For every $\alpha \in E$, the irreducible equation of α over k is of type $X^{p^n} - a = 0$ with some $n \geqq 0$ and $a \in k$.

P. Ins. 4. There exists a set of generators $\{\alpha_i\}_{i \in I}$ of E over k such that each α_i is purely inseparable over k.

To prove the equivalence, assume **P. Ins. 1**. Let $\alpha \in E$. By Theorem 4.1, we conclude that $[k(\alpha):k]_s = 1$. Let $f(X) = \text{Irr}(\alpha, k, X)$. Then f has only one root since

$$[k(\alpha):k]_s$$

is equal to the number of distinct roots of $f(X)$. Let $m = [k(\alpha):k]$. Then $\deg f = m$, and the factorization of f over $k(\alpha)$ is $f(X) = (X - \alpha)^m$. Write $m = p^n r$ where r is an integer prime to p. Then

$$\begin{aligned} f(X) &= (X^{p^n} - \alpha^{p^n})^r \\ &= X^{p^n r} - r\alpha^{p^n} X^{p^n(r-1)} + \text{lower terms.} \end{aligned}$$

Since the coefficients of $f(X)$ lie in k, it follows that

$$r\alpha^{p^n}$$

lies in k, and since $r \neq 0$ (in k), then α^{p^n} lies in k. Let $a = \alpha^{p^n}$. Then α is a root of the polynomial $X^{p^n} - a$, which divides $f(X)$. It follows that $f(X) = X^{p^n} - a$.

Essentially the same argument as the preceding one shows that **P. Ins. 2** implies **P. Ins. 3**. It is trivial that the third condition implies the fourth.

Finally, assume **P. Ins. 4**. Let E be an extension generated by purely inseparable elements α_i $(i \in I)$. Any embedding of E over k maps α_i on a root of

$$f_i(X) = \operatorname{Irr}(\alpha_i, k, X).$$

But $f_i(X)$ divides some polynomial $X^{p^n} - a$, which has only one root. Hence any embedding of E over k is the identity on each α_i, whence the identity on E, and we conclude that $[E:k]_s = 1$, as desired.

An extension satisfying the above four properties will be called **purely inseparable**.

Proposition 6.5. *Purely inseparable extensions form a distinguished class of extensions.*

Proof. The tower theorem is clear from Theorem 4.1, and the lifting property is clear from condition **P. Ins. 4**.

Proposition 6.6. *Let E be an algebraic extension of k. Let E_0 be the compositum of all subfields F of E such that $F \supset k$ and F is separable over k. Then E_0 is separable over k, and E is purely inseparable over E_0.*

Proof. Since separable extensions form a distinguished class, we know that E_0 is separable over k. In fact, E_0 consists of all elements of E which are separable over k. By Proposition 6.1, given $\alpha \in E$ there exists a power of p, say p^n such that α^{p^n} is separable over k. Hence E is purely inseparable over E_0, as was to be shown.

Corollary 6.7. *If an algebraic extension E of k is both separable and purely inseparable, then $E = k$.*

Proof. Obvious.

Corollary 6.8. *Let K be normal over k and let K_0 be its maximal separable subextension. Then K_0 is also normal over k.*

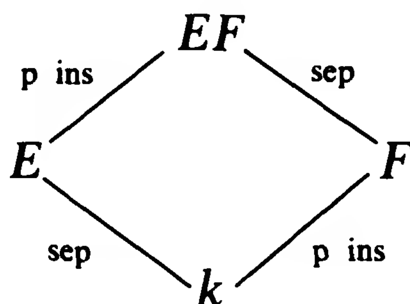
Proof. Let σ be an embedding of K_0 in K^a over k and extend σ to an embedding of K. Then σ is an automorphism of K. Furthermore, σK_0 is separable over k, hence is contained in K_0, since K_0 is the maximal separable subfield. Hence $\sigma K_0 = K_0$, as contended.

Corollary 6.9. *Let E, F be two finite extensions of k, and assume that E/k is separable, F/k is purely inseparable. Assume E, F are subfields of a common field. Then*

$$[EF:F] = [E:k] = [EF:k]_s,$$

$$[EF:E] = [F:k] = [EF:k]_i.$$

Proof. The picture is as follows:



The proof is a trivial juggling of indices, using the corollaries of Proposition 6.1. We leave it as an exercise.

Corollary 6.10. *Let E^p denote the field of all elements x^p, $x \in E$. Let E be a finite extension of k. If $E^pk = E$, then E is separable over k. If E is separable over k, then $E^{p^n}k = E$ for all $n \geqq 1$.*

Proof. Let E_0 be the maximal separable subfield of E. Assume $E^pk = E$. Let $E = k(\alpha_1, \dots, \alpha_n)$. Since E is purely inseparable over E_0 there exists m such that $\alpha_i^{p^m} \in E_0$ for each $i = 1, \dots, n$. Hence $E^{p^m} \subset E_0$. But $E^{p^m}k = E$ whence $E = E_0$ is separable over k. Conversely, assume that E is separable over k. Then E is separable over E^pk. Since E is also purely inseparable over E^pk we conclude that $E = E^pk$. Similarly we get $E = E^{p^n}k$ *for* $n \geqq 1$, as was to be shown.

Proposition 6.6 shows that any algebraic extension can be decomposed into a tower consisting of a maximal separable subextension and a purely inseparable step above it. Usually, one cannot reverse the order of the tower. However, there is an important case when it can be done.

Proposition 6.11. *Let K be normal over k. Let G be its group of automorphisms over k. Let K^G be the fixed field of G (see Chapter* VI, §1). *Then K^G is purely inseparable over k, and K is separable over K^G. If K_0 is the maximal separable subextension of K, then $K = K^G K_0$ and $K_0 \cap K^G = k$.*

Proof. Let $\alpha \in K^G$. Let τ be an embedding of $k(\alpha)$ over k in K^a and extend τ to an embedding of K, which we denote also by τ. Then τ is an automorphism of K because K is normal over k. By definition, $\tau\alpha = \alpha$ and hence τ is the identity on $k(\alpha)$. Hence $[k(\alpha):k]_s = 1$ and α is purely inseparable. Thus K^G is purely inseparable over k. The intersection of K_0

and K^G is both separable and purely inseparable over k, and hence is equal to k.

To prove that K is separable over K^G, assume first that K is finite over k, and hence that G is finite, by Theorem 4.1. Let $\alpha \in K$. Let $\sigma_1, \ldots, \sigma_r$ be a maximal subset of elements of G such that the elements

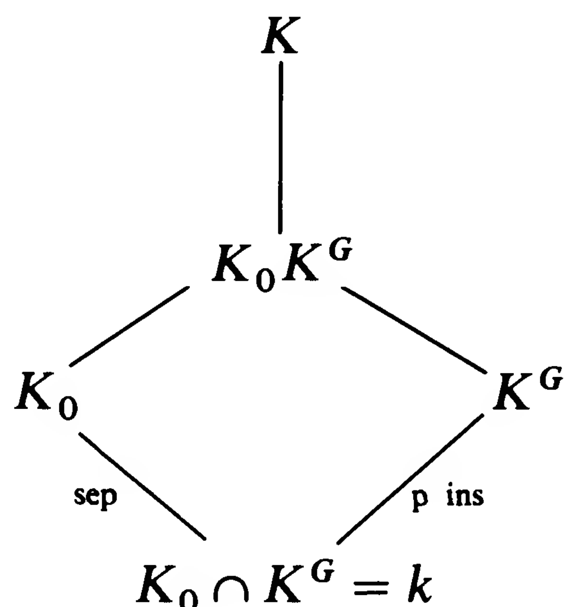
$$\sigma_1\alpha, \ldots, \sigma_r\alpha$$

are distinct, and such that σ_1 is the identity, and α is a root of the polynomial

$$f(X) = \prod_{i=1}^{r} (X - \sigma_i\alpha).$$

For any $\tau \in G$ we note that $f^\tau = f$ because τ permutes the roots. We note that f is separable, and that its coefficients are in the fixed field K^G. Hence α is separable over K^G. The reduction of the infinite case to the finite case is done by observing that every $\alpha \in K$ is contained in some finite normal subextension of K. We leave the details to the reader.

We now have the following picture:



By Proposition 6.6, K is purely inseparable over K_0, hence purely inseparable over K_0K^G. Furthermore, K is separable over K^G, hence separable over K_0K^G. Hence $K = K_0K^G$, thereby proving our proposition.

We see that every normal extension decomposes into a compositum of a purely inseparable and a separable extension. We shall define a Galois extension in the next chapter to be a normal separable extension. Then K_0 is Galois over k and the normal extension is decomposed into a Galois and a purely inseparable extension. The group G is called the **Galois group** of the extension K/k.

A field k is called **perfect** if $k^p = k$. (Every field of characteristic zero is also called perfect.)

Corollary 6.12. *If k is perfect, then every algebraic extension of k is separable, and every algebraic extension of k is perfect.*

Proof. Every finite algebraic extension is contained in a normal extension, and we apply Proposition 6.11 to get what we want.

EXERCISES

1. Let $E = \mathbf{Q}(\alpha)$, where α is a root of the equation

$$\alpha^3 + \alpha^2 + \alpha + 2 = 0.$$

Express $(\alpha^2 + \alpha + 1)(\alpha^2 + \alpha)$ and $(\alpha - 1)^{-1}$ in the form

$$a\alpha^2 + b\alpha + c$$

with $a, b, c \in \mathbf{Q}$.

2. Let $E = F(\alpha)$ where α is algebraic over F, of odd degree. Show that $E = F(\alpha^2)$.

3. Let α and β be two elements which are algebraic over F. Let $f(X) = \mathrm{Irr}(\alpha, F, X)$ and $g(X) = \mathrm{Irr}(\beta, F, X)$. Suppose that $\deg f$ and $\deg g$ are relatively prime. Show that g is irreducible in the polynomial ring $F(\alpha)[X]$.

4. Let α be the real positive fourth root of 2. Find all intermediate fields in the extension $\mathbf{Q}(\alpha)$ of $\mathbf{Q}$.

5. If α is a complex root of $X^6 + X^3 + 1$, find all homomorphisms $\sigma: \mathbf{Q}(\alpha) \to \mathbf{C}$. [*Hint*: The polynomial is a factor of $X^9 - 1$.]

6. Show that $\sqrt{2} + \sqrt{3}$ is algebraic over $\mathbf{Q}$, of degree 4.

7. Let E, F be two finite extensions of a field k, contained in a larger field K. Show that

$$[EF : k] \leqq [E : k][F : k].$$

If $[E : k]$ and $[F : k]$ are relatively prime, show that one has an equality sign in the above relation.

8. Let $f(X) \in k[X]$ be a polynomial of degree n. Let K be its splitting field. Show that $[K:k]$ divides $n!$

9. Find the splitting field of $X^{p^8} - 1$ over the field $\mathbf{Z}/p\mathbf{Z}$.

10. Let α be a real number such that $\alpha^4 = 5$.
    (a) Show that $\mathbf{Q}(i\alpha^2)$ is normal over $\mathbf{Q}$.
    (b) Show that $\mathbf{Q}(\alpha + i\alpha)$ is normal over $\mathbf{Q}(i\alpha^2)$.
    (c) Show that $\mathbf{Q}(\alpha + i\alpha)$ is not normal over $\mathbf{Q}$.

11. Describe the splitting fields of the following polynomials over $\mathbf{Q}$, and find the degree of each such splitting field.
    (a) $X^2 - 2$ (b) $X^2 - 1$
    (c) $X^3 - 2$ (d) $(X^3 - 2)(X^2 - 2)$
    (e) $X^2 + X + 1$ (f) $X^6 + X^3 + 1$
    (g) $X^5 - 7$

12. Let K be a finite field with p^n elements. Show that every element of K has a unique p-th root in K.

13. If the roots of a monic polynomial $f(X) \in k[X]$ in some splitting field are distinct, and form a field, then char $k = p$ and $f(X) = X^{p^n} - X$ for some $n \geqq 1$.

14. Let char $K = p$. Let L be a finite extension of K, and suppose $[L:K]$ prime to p. Show that L is separable over K.

15. Suppose char $K = p$. Let $a \in K$. If a has no p-th root in K, show that $X^{p^n} - a$ is irreducible in $K[X]$ for all positive integers n.

16. Let char $K = p$. Let α be algebraic over K. Show that α is separable if and only if $K(\alpha) = K(\alpha^{p^n})$ for all positive integers n.

17. Prove that the following two properties are equivalent:
    (a) Every algebraic extension of K is separable.
    (b) Either char $K = 0$, or char $K = p$ and every element of K has a p-th root in K.

18. Show that every element of a finite field can be written as a sum of two squares in that field.

19. Let E be an algebraic extension of F. Show that every subring of E which contains F is actually a field. Is this necessarily true if E is not algebraic over F? Prove or give a counterexample.

20. (a) Let $E = F(x)$ where x is transcendental over F. Let $K \neq F$ be a subfield of E which contains F. Show that x is algebraic over K.
    (b) Let $E = F(x)$. Let $y = f(x)/g(x)$ be a rational function, with relatively prime polynomials $f, g \in F[x]$. Let $n = \max(\deg f, \deg g)$. Suppose $n \geqq 1$. Prove that

$$[F(x):F(y)] = n.$$

21. Let $\mathbf{Z}^+$ be the set of positive integers, and A an additive abelian group. Let $f: \mathbf{Z}^+ \to A$ and $g: \mathbf{Z}^+ \to A$ be maps. Suppose that for all n,

$$f(n) = \sum_{d|n} g(d).$$

Let μ be the Möbius function (cf. Exercise 12 of Chapter II). Prove that

$$g(n) = \sum_{d|n} \mu(n/d) f(d).$$

22. Let k be a finite field with q elements. Let $f(X) \in k[X]$ be irreducible. Show that $f(X)$ divides $X^{q^n} - X$ if and only if $\deg f$ divides n. Show the multiplication formula

$$X^{q^n} - X = \prod_{d|n} \prod_{f_d \text{ irr}} f_d(X),$$

where the inner product is over all irreducible polynomials of degree d with leading coefficient 1. Counting degrees, show that

$$q^n = \sum_{d|n} d\psi(d),$$

where $\psi(d)$ is the number of irreducible polynomials of degree d. Invert by

Exercise 21 and find that

$$n\psi(n) = \sum_{d|n} \mu(d)q^{n/d}.$$

23. (a) Let k be a finite field with q elements. Define the **zeta function**

$$Z(t) = (1 - t)^{-1} \prod_p (1 - t^{\deg p})^{-1},$$

where p ranges over all irreducible polynomials $p = p(X)$ in $k[X]$ with leading coefficient 1. Prove that $Z(t)$ is a rational function and determine this rational function.

(b) Let $\pi_q(n)$ be the number of primes p as in (a) of degree $\leqq n$. Prove that

$$\pi_q(m) \sim \frac{q}{q-1}\frac{q^m}{m} \quad \text{for} \quad m \to \infty.$$

Remark. This is the analogue of the prime number theorem in number theory, but it is essentially trivial in the present case, because the Riemann hypothesis is trivially verified. Things get more interesting fast after this case. Consider an equation $y^2 = x^3 + ax + b$ over a finite field $\mathbf{F}_q$ of characteristic $\neq 2$, 3, and having q elements. Assume $-4a^3 - 27b^2 \neq 0$, in which case the curve defined by this equation is called an **elliptic curve**. Define N_n by

$N_n - 1$ = number of points (x, y) satisfying the above equation with $x, y \in \mathbf{F}_{q^n}$ (the extension of $\mathbf{F}_q$ of degree n).

Define the **zeta function** $Z(t)$ to be the unique rational function such that $Z(0) = 1$ and

$$Z'/Z(t) = \sum N_n t^{n-1}.$$

A famous theorem of Hasse asserts that $Z(t)$ is a rational function of the form

$$Z(t) = \frac{(1 - \alpha t)(1 - \bar{\alpha} t)}{(1 - t)(1 - qt)},$$

where α is an imaginary quadratic number (not real, quadratic over $\mathbf{Q}$), $\bar{\alpha}$ is its complex conjugate, and $\alpha\bar{\alpha} = q$, so $|\alpha| = q^{1/2}$. See Hasse, "Abstrakte Bergrundung der komplexen Multiplikation und Riemannsche Vermutung in Funktionenkörpern," *Abh. Math. Sem. Univ. Hamburg* **10** (1934) pp. 325–348.

24. Let k be a field of characteristic p and let t, u be algebraically independent over k. Prove the following:
(a) $k(t, u)$ has degree p^2 over $k(t^p, u^p)$.
(b) There exist infinitely many extensions between $k(t, u)$ and $k(t^p, u^p)$.

25. Let E be a finite extension of k and let $p^r = [E:k]_i$. We assume that the characteristic is $p > 0$. Assume that there is no exponent p^s with $s < r$ such that $E^{p^s}k$ is separable over k (i.e., such that α^{p^s} is separable over k for each α in E). Show that E can be generated by one element over k. [*Hint*: Assume first that E is purely inseparable.]

26. Let k be a field, $f(X)$ an irreducible polynomial in $k[X]$, and let K be a finite normal extension of k. If g, h are monic irreducible factors of $f(X)$ in $K[X]$, show that there exists an automorphism σ of K over k such that $g = h^\sigma$. Give an example when this conclusion is not valid if K is not normal over k.

27. Let $x_1, \ldots, x_n$ be algebraically independent over a field k. Let y be algebraic over $k(x) = k(x_1, \ldots, x_n)$. Let $P(X_{n+1})$ be the irreducible polynomial of y over $k(x)$. Let $\varphi(x)$ be the least common multiple of the denominators of the coefficients of P. Then the coefficients of $\varphi(x)P$ are elements of $k[x]$. Show that the polynomial

$$f(X_1, \ldots, X_{n+1}) = \varphi(X_1, \ldots, X_n)P(X_{n+1})$$

is irreducible over k, as a polynomial in $n + 1$ variables.

Conversely, let $f(X_1, \ldots, X_{n+1})$ be an irreducible polynomial over k. Let $x_1, \ldots, x_n$ be algebraically independent over k. Show that

$$f(x_1, \ldots, x_n, X_{n+1})$$

is irreducible over $k(x_1, \ldots, x_n)$.

If f is a polynomial in n variables, and $(b) = (b_1, \ldots, b_n)$ is an n-tuple of elements such that $f(b) = 0$, then we say that (b) is a **zero** of f. We say that (b) is **non-trivial** if not all coordinates b_i are equal to 0.

28. Let $f(X_1, \ldots, X_n)$ be a homogeneous polynomial of degree 2 (resp. 3) over a field k. Show that if f has a non-trivial zero in an extension of odd degree (resp. degree 2) over k, then f has a non-trivial zero in k.

29. Let $f(X, Y)$ be an irreducible polynomial in two variables over a field k. Let t be transcendental over k, and assume that there exist integers $m, n \neq 0$ and elements $a, b \in k$, $ab \neq 0$, such that $f(at^n, bt^m) = 0$. Show that after inverting possibly X or Y, and up to a constant factor, f is of type

$$X^m Y^n - c$$

with some $c \in k$.

The answer to the following exercise is not known.

30. **(Artin conjecture).** Let f be a homogeneous polynomial of degree d in n variables, with rational coefficients. If $n > d$, show that there exists a root of unity ζ, and elements

$$x_1, \ldots, x_n \in \mathbf{Q}[\zeta]$$

not all 0 such that $f(x_1, \ldots, x_n) = 0$.

31. **Difference equations.** Let $u_1, \ldots, u_d$ be elements of a field K. We want to solve for infinite vectors $(x_0, x_1, \ldots, x_n, \ldots)$ satisfying

$$(*) \qquad x_n = u_1 x_{n-1} + \cdots + u_d x_{n-d} \qquad \text{for} \quad n \geqq d.$$

Define the **characteristic polynomial** of the system to be

$$X^d - (u_1 X^{d-1} + \cdots + u_d) = f(X).$$

Suppose α is a root of f.

(a) Show that $x_n = \alpha^n$ $(n \geqq 0)$ is a solution of (*).

(b) Show that the set of solutions of (*) is a vector space of dimension d.

(c) Assume that the characteristic polynomial has d distinct roots $\alpha_1, \ldots, \alpha_d$. Show that the solutions $(\alpha_1^n), \ldots, (\alpha_d^n)$ form a basis for the space of solutions.

(d) Let $x_n = b_1\alpha_1^n + \cdots + b_d\alpha_d^n$ for $n \geqq 0$, show how to solve for $b_1, \ldots, b_d$ in terms of $\alpha_1, \ldots, \alpha_d$ and $x_0, \ldots, x_{d-1}$. (Use the Vandermonde determinant.)

(e) Under the conditions of (d), let $F(T) = \sum x_n T^n$. Show that $F(T)$ represents a rational function, and give its partial fraction decomposition.

32. Let $d = 2$ for simplicity. Given $a_0, a_1, u, v, w, t \in K$, we want to find the solutions of the system

$$a_n = ua_{n-1} - vta_{n-2} - t^n w \qquad \text{for} \quad n \geqq 2.$$

Let α_1, α_2 be the root of the characteristic polynomial, that is

$$1 - uX + vtX^2 = (1 - \alpha_1 X)(1 - \alpha_2 X).$$

Assume that α_1, α_2 *are distinct, and also distinct from* t. *Let*

$$F(X) = \sum_{n=0}^{\infty} a_n X^n.$$

(a) Show that there exist elements A, B, C of K such that

$$F(X) = \frac{A}{1 - \alpha_1 X} + \frac{B}{1 - \alpha_2 X} + \frac{C}{1 - tX}.$$

(b) Show that there is a unique solution to the difference equation given by

$$a_n = A\alpha_1^n + B\alpha_2^n + Ct^n \qquad \text{for} \quad n \geqq 0.$$

(To see an application of this formalism to modular forms, as in the work of Manin, Mazur, and Swinnerton-Dyer, cf. my *Introduction to Modular Forms*, Springer-Verlag, New York, 1976, Chapter XII, §2.)

33. Let R be a ring which we assume entire for simplicity. Let

$$g(T) = T^d - a_{d-1}T^{d-1} - \cdots - a_0$$

be a polynomial in $R[T]$, and consider the equation

$$T^d = a_0 + a_1 T + \cdots + a_{d-1}T^{d-1}.$$

Let x be a root of $g(T)$.

(a) For any integer $n \geqq d$ there is a relation

$$x^n = a_{0,n} + a_{1,n}x + \cdots + a_{d-1,n}x^{d-1}$$

with coefficients $a_{i,j}$ in $\mathbf{Z}[a_0, \ldots, a_{d-1}] \subset R$.

(b) Let $F(T) \in R[T]$ be a polynomial. Then

$$F(x) = a_0(F) + a_1(F)x + \cdots + a_{d-1}(F)x^{d-1}$$

where the coefficients $a_i(F)$ lie in R and depend linearly on F.

(c) Let the Vandermonde determinant be

$$V(x_1, \dots, x_d) = \begin{vmatrix} 1 & x_1 & \cdots & x_1^{d-1} \\ 1 & x_2 & \cdots & x_2^{d-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_d & \cdots & x_d^{d-1} \end{vmatrix} = \prod_{i<j} (x_j - x_i).$$

Suppose that the equation $g(T) = 0$ has d roots and that there is a factorization

$$g(T) = \prod_{i=1}^{d} (T - x_i).$$

Substituting x_i for x with $i = 1, \dots, d$ and using Cramer's rule on the resulting system of linear equations, yields

$$\Delta a_j(F) = \Delta_j(F)$$

where Δ is the Vandermonde determinant, and $\Delta_j(F)$ is obtained by replacing the j-th column by ${}^t(F(x_1), \dots, F(x_d))$, so

$$\Delta_j(F) = \begin{vmatrix} 1 & x_1 & \cdots & F(x_1) & \cdots & x_1^{d-1} \\ 1 & x_2 & \cdots & F(x_2) & \cdots & x_2^{d-1} \\ \vdots & \vdots & & \vdots & & \vdots \\ 1 & x_d & \cdots & F(x_d) & \cdots & x_d^{d-1} \end{vmatrix}$$

If $\Delta \neq 0$ then we can write

$$a_j(F) = \Delta_j(F)/\Delta.$$

Remark. If $F(T)$ is a power series in $R[[T]]$ and if R is a complete local ring, with $x_1, \dots, x_d$ in the maximal ideal, and $x = x_i$ for some i, then we can evaluate $F(x)$ because the series converges. The above formula for the coefficients $a_j(F)$ remains valid.

34. Let $x_1, \dots, x_d$ be independent variables, and let A be the ring

$$\mathbf{Q}[[x_1, \dots, x_d]][T]/\prod_{i=1}^{d} (T - x_i).$$

Substituting some x_i for T induces a natural homomorphism φ_i of A onto

$$\mathbf{Q}[[z_1, \dots, x_d]] = R,$$

and the map $z \mapsto (\varphi_1(z), \dots, \varphi_d(z))$ gives an embedding of A into the product of R with itself d times.

Let k be an integer, and consider the formal power series

$$F(T) = e^{kT} \prod_{i=1}^{d} \frac{(T - x_i)e^{T-x_i}}{e^{T-x_i} - 1} = e^{kT} \prod_{i=1}^{d} h(T - x_i)$$

where $h(t) = te^t/(e^t - 1)$. It is a formal power series in T, $T - x_1, \dots, T - x_d$. Under substitution of some x_j for T it becomes a power series in x_j and $x_j - x_i$, and thus converges in $\mathbf{Q}[[x_1, \dots, x_d]]$.

(a) Verify that

$$F(T) \equiv a_0(F) + \cdots + a_{d-1}(F)T^{d-1} \bmod \prod_{i=1}^{d} (T - x_i)$$

where $a_0(F), \ldots, a_{d-1}(F) \in \mathbf{Q}[[x_1, \ldots, x_d]]$, and that the formula given in the preceding exercise for these coefficients in terms of Vandermonde determinants is valid.

(b) Show that $a_{d-1}(F) = 0$ if $-(d-1) \leqq k < 0$ and $a_{d-1}(F) = 1$ if $k = 0$.

Remark. The assertion in (a) is a simple limit. The assertion in (b) is a fact which has been used in the proof of the Hirzebruch–Grothendieck–Riemann–Roch theorem and as far as I know there was no simple known proof until Roger Howe pointed out that it could be done by the formula of the preceding exercise as follows. We have

$$V(x_1, \ldots, x_n)a_{d-1}(F) = \begin{vmatrix} 1 & x_1 & \cdots & x_1^{d-2} & F(x_1) \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & x_d & \cdots & x_d^{d-2} & F(x_d) \end{vmatrix}.$$

Furthermore,

$$F(x_j) = e^{kx_j} \prod_{n \neq j} \frac{(x_j - x_n)e^{x_j - x_n}}{e^{x_j - x_n} - 1}.$$

We use the inductive relation of Vandermonde determinants

$$V(x_1, \ldots, x_d) = V(x_1, \ldots, \hat{x}_j, \ldots, x_d)(-1)^{d-j} \prod_{n \neq j} (x_j - x_n).$$

We expand the determinant for $a_{d-1}(F)$ according to the last column to get

$$a_{d-1}(F) = \sum_{j=1}^{d} e^{(k+d-1)x_j} \prod_{n \neq j} \frac{1}{e^{x_j} - e^{x_n}}.$$

Using the inductive relation backward, and replacing x_i by e^{x_i} which we denote by y_i for typographical reasons, we get

$$V(y_1, \ldots, y_d)a_{d-1}(F) = \begin{vmatrix} 1 & y_1 & \cdots & y_1^{d-2} & y_1^{k+d-1} \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & y_d & \cdots & y_d^{d-2} & y_d^{k+d-1} \end{vmatrix}$$

If $k \neq 0$ then two columns on the right are the same, so the determinant is 0. If $k = 0$ then we get the Vandermonde determinant on the right, so $a_{d-1}(F) = 1$. This proves the desired value.

CHAPTER VI

Galois Theory

This chapter contains the core of Galois theory. We study the group of automorphisms of a finite (and sometimes infinite) Galois extension at length, and give examples, such as cyclotomic extensions, abelian extensions, and even non-abelian ones, leading into the study of matrix representations of the Galois group and their classifications. We shall mention a number of fundamental unsolved problems, the most notable of which is whether given a finite group G, there exists a Galois extension of **Q** having this group as Galois group. Three surveys give recent points of view on those questions and sizeable bibliographies:

B. MATZAT, *Konstruktive Galoistheorie*, Springer Lecture Notes **1284**, 1987

B. MATZAT, Über das Umkehrproblem der Galoisschen Theorie, *Jahrsbericht Deutsch. Mat.-Verein.* **90** (1988), pp. 155–183

J. P. SERRE, *Topics in Galois theory*, course at Harvard, 1989, Jones and Bartlett, Boston 1992

More specific references will be given in the text at the appropriate moment concerning this problem and the problem of determining Galois groups over specific fields, especially the rational numbers.

§1. GALOIS EXTENSIONS

Let K be a field and let G be a group of automorphisms of K. We denote by K^G the subset of K consisting of all elements $x \in K$ such that $x^\sigma = x$ for all $\sigma \in G$. It is also called the **fixed field** of G. It is a field because if $x, y \in K^G$ then

$$(x + y)^\sigma = x^\sigma + y^\sigma = x + y$$

for all $\sigma \in G$, and similarly, one verifies that K is closed under multiplication, subtraction, and multiplicative inverse. Furthermore, K^G contains 0 and 1, hence contains the prime field.

An algebraic extension K of a field k is called **Galois** if it is normal and separable. We consider K as embedded in an algebraic closure. The group of automorphisms of K over k is called the **Galois group** of K over k, and is denoted by $G(K/k)$, $G_{K/k}$, $\text{Gal}(K/k)$, or simply G. It coincides with the set of embeddings of K in K^a over k.

For the convenience of the reader, we shall now state the main result of the Galois theory for finite Galois extensions.

Theorem 1.1. *Let K be a finite Galois extension of k, with Galois group G. There is a bijection between the set of subfields E of K containing k, and the set of subgroups H of G, given by $E = K^H$. The field E is Galois over k if and only if H is normal in G, and if that is the case, then the map $\sigma \mapsto \sigma | E$ induces an isomorphism of G/H onto the Galois group of E over k.*

We shall give the proofs step by step, and as far as possible, we give them for infinite extensions.

Theorem 1.2. *Let K be a Galois extension of k. Let G be its Galois group. Then $k = K^G$. If F is an intermediate field, $k \subset F \subset K$, then K is Galois over F. The map*

$$F \mapsto G(K/F)$$

from the set of intermediate fields into the set of subgroups of G is injective.

Proof. Let $\alpha \in K^G$. Let σ be any embedding of $k(\alpha)$ in K^a, inducing the identity on k. Extend σ to an embedding of K into K^a, and call this extension σ also. Then σ is an automorphism of K over k, hence is an element of G. By assumption, σ leaves α fixed. Therefore

$$[k(\alpha):k]_s = 1.$$

Since α is separable over k, we have $k(\alpha) = k$ and α is an element of k. This proves our first assertion.

Let F be an intermediate field. Then K is normal and separable over F by Theorem 3.4 and Theorem 4.5 of Chapter V. Hence K is Galois over F. If $H = G(K/F)$ then by what we proved above we conclude that $F = K^H$. If F, F' are intermediate fields, and $H = G(K/F)$, $H' = G(K/F')$, then

$$F = K^H \quad \text{and} \quad F' = K^{H'}.$$

If $H = H'$ we conclude that $F = F'$, whence our map

$$F \mapsto G(K/F)$$

is injective, thereby proving our theorem.

We shall sometimes call the group $G(K/F)$ of an intermediate field the group **associated** with F. We say that a subgroup H of G **belongs** to an intermediate field F if $H = G(K/F)$.

Corollary 1.3. *Let K/k be Galois with group G. Let F, F' be two intermediate fields, and let H, H' be the subgroups of G belonging to F, F' respectively. Then $H \cap H'$ belongs to FF'.*

Proof. Every element of $H \cap H'$ leaves FF' fixed, and every element of G which leaves FF' fixed also leaves F and F' fixed and hence lies in $H \cap H'$. This proves our assertion.

Corollary 1.4. *Let the notation be as in Corollary 1.3. The fixed field of the smallest subgroup of G containing H, H' is $F \cap F'$.*

Proof. Obvious.

Corollary 1.5. *Let the notation be as in Corollary 1.3. Then $F \subset F'$ if and only if $H' \subset H$.*

Proof. If $F \subset F'$ and $\sigma \in H'$ leaves F' fixed then σ leaves F fixed, so σ lies in H. Conversely, if $H' \subset H$ then the fixed field of H is contained in the fixed field of H', so $F \subset F'$.

Corollary 1.6. *Let E be a finite separable extension of a field k. Let K be the smallest normal extension of k containing E. Then K is finite Galois over k. There is only a finite number of intermediate fields F such that $k \subset F \subset E$.*

Proof. We know that K is normal and separable, and K is finite over k since we saw that it is the finite compositum of the finite number of conjugates of E. The Galois group of K/k has only a finite number of subgroups. Hence there is only a finite number of subfields of K containing k, whence *a fortiori* a finite number of subfields of E containing k.

Of course, the last assertion of Corollary 1.6 has been proved in the preceding chapter, but we get another proof here from another point of view.

Lemma 1.7. *Let E be an algebraic separable extension of k. Assume that there is an integer $n \geqq 1$ such that every element α of E is of degree $\leqq n$ over k. Then E is finite over k and $[E:k] \leqq n$.*

Proof. Let α be an element of E such that the degree $[k(\alpha):k]$ is maximal, say $m \leqq n$. We contend that $k(\alpha) = E$. If this is not true, then there exists an element $\beta \in E$ such that $\beta \notin k(\alpha)$, and by the primitive element theorem, there exists an element $\gamma \in k(\alpha, \beta)$ such that $k(\alpha, \beta) = k(\gamma)$. But from the tower

$$k \subset k(\alpha) \subset k(\alpha, \beta)$$

we see that $[k(\alpha, \beta):k] > m$ whence γ has degree $> m$ over k, contradiction.

Theorem 1.8. (Artin). *Let K be a field and let G be a finite group of automorphisms of K, of order n. Let $k = K^G$ be the fixed field. Then K is a finite Galois extension of k, and its Galois group is G. We have $[K : k] = n$.*

Proof. Let $\alpha \in K$ and let $\sigma_1, \dots, \sigma_r$ be a maximal set of elements of G such that $\sigma_1\alpha, \dots, \sigma_r\alpha$ are distinct. If $\tau \in G$ then $(\tau\sigma_1\alpha, \dots, \tau\sigma_r\alpha)$ differs from $(\sigma_1\alpha, \dots, \sigma_r\alpha)$ by a permutation, because τ is injective, and every $\tau\sigma_i\alpha$ is among the set $\{\sigma_1\alpha, \dots, \sigma_r\alpha\}$; otherwise this set is not maximal. Hence α is a root of the polynomial

$$f(X) = \prod_{i=1}^{r} (X - \sigma_i\alpha),$$

and for any $\tau \in G$, $f^\tau = f$. Hence the coefficients of f lie in $K^G = k$. Furthermore, f is separable. Hence every element α of K is a root of a separable polynomial of degree $\leqq n$ with coefficients in k. Furthermore, this polynomial splits in linear factors in K. Hence K is separable over k, is normal over k, hence Galois over k. By Lemma 1.7, we have $[K : k] \leqq n$. The Galois group of K over k has order $\leqq [K:k]$ (by Theorem 4.1 of Chapter V), and hence G must be the full Galois group. This proves all our assertions.

Corollary 1.9. *Let K be a finite Galois extension of k and let G be its Galois group. Then every subgroup of G belongs to some subfield F such that $k \subset F \subset K$.*

Proof. Let H be a subgroup of G and let $F = K^H$. By Artin's theorem we know that K is Galois over F with group H.

Remark. When K is an infinite Galois extension of k, then the preceding corollary is not true any more. This shows that some counting argument must be used in the proof of the finite case. In the present treatment, we have used an old-fashioned argument. The reader can look up Artin's own proof in his book *Galois Theory*. In the infinite case, one defines the Krull topology on the Galois group G (cf. exercises 43–45), and G becomes a compact totally disconnected group. The subgroups which belong to the intermediate fields are the *closed* subgroups. The reader may disregard the infinite case entirely throughout our discussions without impairing understanding. The proofs in the infinite case are usually identical with those in the finite case.

The notions of a Galois extension and a Galois group are defined completely algebraically. Hence they behave formally under isomorphisms the way one expects from objects in any category. We describe this behavior more explicitly in the present case.

Let K be a Galois extension of k. Let

$$\lambda : K \to \lambda K$$

be an isomorphism. Then λK is a Galois extension of λk.

$$\begin{array}{ccc} K & \xrightarrow{\lambda} & \lambda K \\ | & & | \\ k & \xrightarrow[\lambda]{} & \lambda k \end{array}$$

Let G be the Galois group of K over k. Then the map

$$\sigma \mapsto \lambda \circ \sigma \circ \lambda^{-1}$$

gives a homomorphism of G into the Galois group of λK over λk, whose inverse is given by

$$\lambda^{-1} \circ \tau \circ \lambda \leftarrow\!\dashv \tau.$$

Hence $G(\lambda K/\lambda k)$ is isomorphic to $G(K/k)$ under the above map. We may write

$$G(\lambda K/\lambda k)^{\lambda} = G(K/k)$$

or

$$G(\lambda K/\lambda k) = \lambda G(K/k)\lambda^{-1},$$

where the exponent λ is "conjugation,"

$$\sigma^{\lambda} = \lambda^{-1} \circ \sigma \circ \lambda.$$

There is no avoiding the contravariance if we wish to preserve the rule

$$(\sigma^{\lambda})^{\omega} = \sigma^{\lambda\omega}$$

when we compose mappings λ and ω.

In particular, let F be an intermediate field, $k \subset F \subset K$, and let $\lambda : F \to \lambda F$ be an embedding of F in K, which we assume is extended to an automorphism of K. Then $\lambda K = K$. Hence

$$G(K/\lambda F)^{\lambda} = G(K/F)$$

and

$$G(K/\lambda F) = \lambda G(K/F)\lambda^{-1}.$$

Theorem 1.10. *Let K be a Galois extension of k with group G. Let F be a subfield, $k \subset F \subset K$, and let $H = G(K/F)$. Then F is normal over k if and only if H is normal in G. If F is normal over k, then the restriction map $\sigma \mapsto \sigma | F$*

is a homomorphism of G onto the Galois group of F over k, whose kernel is H. We thus have $G(F/k) \approx G/H$.

Proof. Assume F is normal over k, and let G' be its Galois group. The restriction map $\sigma \to \sigma|F$ maps G into G', and by definition, its kernel is H. Hence H is normal in G. Furthermore, any element $\tau \in G'$ extends to an embedding of K in K^a, which must be an automorphism of K, so the restriction map is surjective. This proves the last statement. Finally, assume that F is not normal over k. Then there exists an embedding λ of F in K over k which is not an automorphism, i.e. $\lambda F \neq F$. Extend λ to an automorphism of K over k. The Galois groups $G(K/\lambda F)$ and $G(K/F)$ are conjugate, and they belong to distinct subfields, hence cannot be equal. Hence H is not normal in G.

A Galois extension K/k is said to be **abelian** (resp. **cyclic**) if its Galois group G is abelian (resp. cyclic).

Corollary 1.11. *Let K/k be abelian (resp. cyclic). If F is an intermediate field, $k \subset F \subset K$, then F is Galois over k and abelian (resp. cyclic).*

Proof. This follows at once from the fact that a subgroup of an abelian group is normal, and a factor group of an abelian (resp. cyclic) group is abelian (resp. cyclic).

Theorem 1.12. *Let K be a Galois extension of k, let F be an arbitrary extension and assume that K, F are subfields of some other field. Then KF is Galois over F, and K is Galois over $K \cap F$. Let H be the Galois group of KF over F, and G the Galois group of K over k. If $\sigma \in H$ then the restriction of σ to K is in G, and the map*

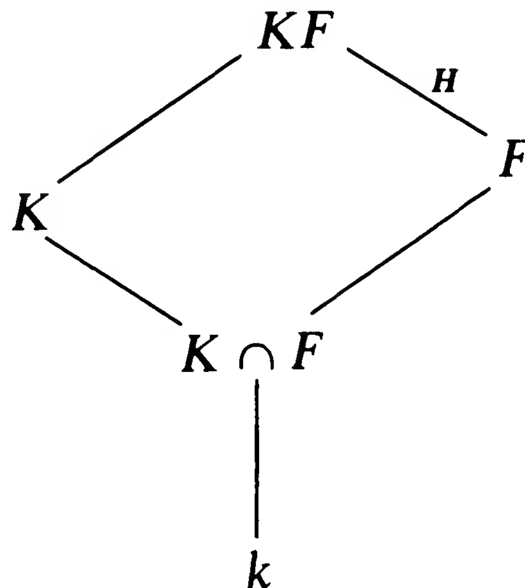
$$\sigma \mapsto \sigma|K$$

gives an isomorphism of H on the Galois group of K over $K \cap F$.

Proof. Let $\sigma \in H$. The restriction of σ to K is an embedding of K over k, whence an element of G since K is normal over k. The map $\sigma \mapsto \sigma|K$ is clearly a homomorphism. If $\sigma|K$ is the identity, then σ must be the identity of KF (since every element of KF can be expressed as a combination of sums, products, and quotients of elements in K and F). Hence our homomorphism $\sigma \mapsto \sigma|K$ is injective. Let H' be its image. Then H' leaves $K \cap F$ fixed, and conversely, if an element $\alpha \in K$ is fixed under H', we see that α is also fixed under H, whence $\alpha \in F$ and $\alpha \in K \cap F$. Therefore $K \cap F$ is the fixed field. If K is finite over k, or even KF finite over F, then by Theorem 1.8, we know that H' is the Galois group of K over $K \cap F$, and the theorem is proved in that case.

(In the infinite case, one must add the remark that for the Krull topology, our map $\sigma \mapsto \sigma|K$ is continuous, whence its image is closed since H is compact. See Theorem 14.1; Chapter I, Theorem 10.1; and Exercise 43.)

The diagram illustrating Theorem 1.12 is as follows:



It is suggestive to think of the opposite sides of a parallelogram as being equal.

Corollary 1.13. *Let K be a finite Galois extension of k. Let F be an arbitrary extension of k. Then $[KF : F]$ divides $[K : k]$.*

Proof. Notation being as above, we know that the order of H divides the order of G, so our assertion follows.

Warning. The assertion of the corollary is not usually valid if K is not Galois over k. For instance, let $\alpha = \sqrt[3]{2}$ be the real cube root of 2, let ζ be a cube root of 1, $\zeta \neq 1$, say

$$\zeta = \frac{-1 + \sqrt{-3}}{2},$$

and let $\beta = \zeta\alpha$. Let $E = \mathbf{Q}(\beta)$. Since β is complex and α real, we have

$$\mathbf{Q}(\beta) \neq \mathbf{Q}(\alpha).$$

Let $F = \mathbf{Q}(\alpha)$. Then $E \cap F$ is a subfield of E whose degree over $\mathbf{Q}$ divides 3. Hence this degree is 3 or 1, and must be 1 since $E \neq F$. But

$$EF = \mathbf{Q}(\alpha, \beta) = \mathbf{Q}(\alpha, \zeta) = \mathbf{Q}(\alpha, \sqrt{-3}).$$

Hence EF has degree 2 over F.

Theorem 1.14. *Let K_1 and K_2 be Galois extensions of a field k, with Galois groups G_1 and G_2 respectively. Assume K_1, K_2 are subfields of some field. Then K_1K_2 is Galois over k. Let G be its Galois group. Map $G \to G_1 \times G_2$ by restriction, namely*

$$\sigma \mapsto (\sigma|K_1, \sigma|K_2).$$

This map is injective. If $K_1 \cap K_2 = k$ then the map is an isomorphism.

Proof. Normality and separability are preserved in taking the compositum of two fields, so K_1K_2 is Galois over k. Our map is obviously a homomorphism of G into $G_1 \times G_2$. If an element $\sigma \in G$ induces the identity on K_1 and K_2 then it induces the identity on their compositum, so our map is injective. Assume that $K_1 \cap K_2 = k$. According to Theorem 1.12, given an element $\sigma_1 \in G_1$ there exists an element σ of the Galois group of K_1K_2 over K_2 which induces σ_1 on K_1. This σ is *a fortiori* in G, and induces the identity on K_2. Hence $G_1 \times \{e_2\}$ is contained in the image of our homomorphism (where e_2 is the unit element of G_2). Similarly, $\{e_1\} \times G_2$ is contained in this image. Hence their product is contained in the image, and their product is precisely $G_1 \times G_2$. This proves Theorem 1.14.

$$\begin{array}{ccc} & K_1K_2 & \\ \diagup & & \diagdown \\ K_1 & & K_2 \\ \diagdown & & \diagup \\ & K_1 \cap K_2 & \\ & | & \\ & k & \end{array}$$

Corollary 1.15. *Let $K_1, \dots, K_n$ be Galois extensions of k with Galois groups $G_1, \dots, G_n$. Assume that $K_{i+1} \cap (K_1 \cdots K_i) = k$ for each $i = 1, \dots, n-1$. Then the Galois group of $K_1 \cdots K_n$ is isomorphic to the product $G_1 \times \cdots \times G_n$ in the natural way.*

Proof. Induction.

Corollary 1.16. *Let K be a finite Galois extension of k with group G, and assume that G can be written as a direct product $G = G_1 \times \cdots \times G_n$. Let K_i be the fixed field of*

$$G_1 \times \cdots \times \{1\} \times \cdots \times G_n$$

where the group with 1 element occurs in the i-th place. Then K_i is Galois over k, and $K_{i+1} \cap (K_1 \cdots K_i) = k$. Furthermore $K = K_1 \cdots K_n$.

Proof. By Corollary 1.3, the compositum of all K_i belongs to the intersection of their corresponding groups, which is clearly the identity. Hence the compositum is equal to K. Each factor of G is normal in G, so K_i is Galois over k. By Corollary 1.4, the intersection of normal extensions belongs to the product of their Galois groups, and it is then clear that $K_{i+1} \cap (K_1 \cdots K_i) = k$.

Theorem 1.17. *Assume all fields contained in some common field.*

(i) *If K, L are abelian over k, so is the composite KL.*

(ii) *If K is abelian over k and E is any extension of k, then KE is abelian over E.*

(iii) *If K is abelian over k and $K \supset E \supset k$ where E is an intermediate field, then E is abelian over k and K is abelian over E.*

Proof. Immediate from Theorems 1.12 and 1.14.

If k is a field, the composite of all abelian extensions of k in a given algebraic closure k^a is called the **maximum abelian extension** of k, and is denoted by k^{ab}.

Remark on notation. We have used systematically the notation:

k^a = algebraic closure of k;

k^s = separable closure of k;

k^{ab} = abelian closure of k = maximal abelian extension.

We have replaced other people's notation $\bar{k}$ (and mine as well in the first edition) with k^a in order to make the notation functorial with respect to the ideas.

§2. EXAMPLES AND APPLICATIONS

Let k be a field and $f(X)$ a separable polynomial of degree $\geqq 1$ in $k[X]$. Let

$$f(X) = (X - \alpha_1) \cdots (X - \alpha_n)$$

be its factorization in a splitting field K over k. Let G be the Galois group of K over k. We call G the **Galois group** of f over k. Then the elements of G permute the roots of f. Thus we have an injective homomorphism of G into the symmetric group S_n on n elements. Not every permutation need be given by an element of G. We shall discuss examples below.

Example 1. Quadratic extensions. Let k be a field and $a \in k$. If a is not a square in k, then the polynomial $X^2 - a$ has no root in k and is therefore irreducible. Assume char $k \neq 2$. Then the polynomial is separable (because $2 \neq 0$), and if α is a root, then $k(\alpha)$ is the splitting field, is Galois, and its Galois group is cyclic of order 2.

Conversely, given an extension K of k of degree 2, there exists $a \in k$ such that $K = k(\alpha)$ and $\alpha^2 = a$. This comes from completing the square and the quadratic formula as in elementary school. The formula is valid as long as the characteristic of k is $\neq 2$.

Example 2. Cubic extensions. Let k be a field of characteristic $\neq 2$ or 3. Let

$$f(X) = X^3 + aX + b.$$

Any polynomial of degree 3 can be brought into this form by completing the cube. Assume that f has no root in k. Then f is irreducible because any factorization must have a factor of degree 1. Let α be a root of $f(X)$. Then

$$[k(\alpha): k] = 3.$$

Let K be the splitting field. Since char $k \neq 2, 3$, f is separable. Let G be the Galois group. Then G has order 3 or 6 since G is a subgroup of the symmetric group S_3. In the second case, $k(\alpha)$ is not normal over k.

There is an easy way to test whether the Galois group is the full symmetric group. We consider the discriminant. If $\alpha_1, \alpha_2, \alpha_3$ are the distinct roots of $f(X)$, we let

$$\delta = (\alpha_1 - \alpha_2)(\alpha_2 - \alpha_3)(\alpha_1 - \alpha_3) \quad \text{and} \quad \Delta = \delta^2.$$

If G is the Galois group and $\sigma \in G$ then $\sigma(\delta) = \pm\delta$. Hence σ leaves Δ fixed. Thus Δ is in the ground field k, and in Chapter IV, §6, we have seen that

$$\Delta = -4a^3 - 27b^2.$$

The set of σ in G which leave δ fixed is precisely the set of even permutations. Thus G is the symmetric group if and only if Δ is not a square in k. We may summarize the above remarks as follows.

Let $f(X)$ be a cubic polynomial in $k[X]$, and assume char $k \neq 2, 3$. *Then:*

(a) *f is irreducible over k if and only if f has no root in k.*

(b) *Assume f irreducible. Then the Galois group of f is S_3 if and only if the discriminant of f is not a square in k. If the discriminant is a square, then the Galois group is cyclic of order* 3, *equal to the alternating group A_3 as a permutation of the roots of f.*

For instance, consider

$$f(X) = X^3 - X + 1$$

over the rational numbers. Any rational root must be 1 or -1, and so $f(X)$ is irreducible over $\mathbf{Q}$. The discriminant is -23, and is not a square. Hence the Galois group is the symmetric group. The splitting field contains a subfield of degree 2, namely $k(\delta) = k(\sqrt{\Delta})$.

On the other hand, let $f(X) = X^3 - 3X + 1$. Then f has no root in $\mathbf{Z}$, whence no root in $\mathbf{Q}$, so f is irreducible. The discriminant is 81, which is a square, so the Galois group is cyclic of order 3.

Example 3. We consider the polynomial $f(X) = X^4 - 2$ over the rationals $\mathbf{Q}$. It is irreducible by Eisenstein's criterion. Let α be a real root.

Let $i = \sqrt{-1}$. Then $\pm\alpha$ and $\pm i\alpha$ are the four roots of $f(X)$, and

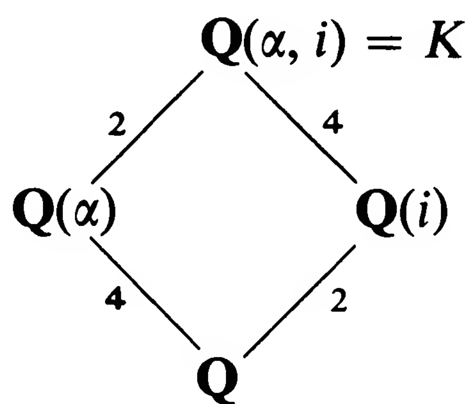
$$[\mathbf{Q}(\alpha) : \mathbf{Q}] = 4.$$

Hence the splitting field of $f(X)$ is

$$K = \mathbf{Q}(\alpha, i).$$

The field $\mathbf{Q}(\alpha) \cap \mathbf{Q}(i)$ has degree 1 or 2 over $\mathbf{Q}$. The degree cannot be 2 otherwise $i \in \mathbf{Q}(\alpha)$, which is impossible since α is real. Hence the degree is 1. Hence i has degree 2 over $\mathbf{Q}(\alpha)$ and therefore $[K : \mathbf{Q}] = 8$. The Galois group of $f(X)$ has order 8.

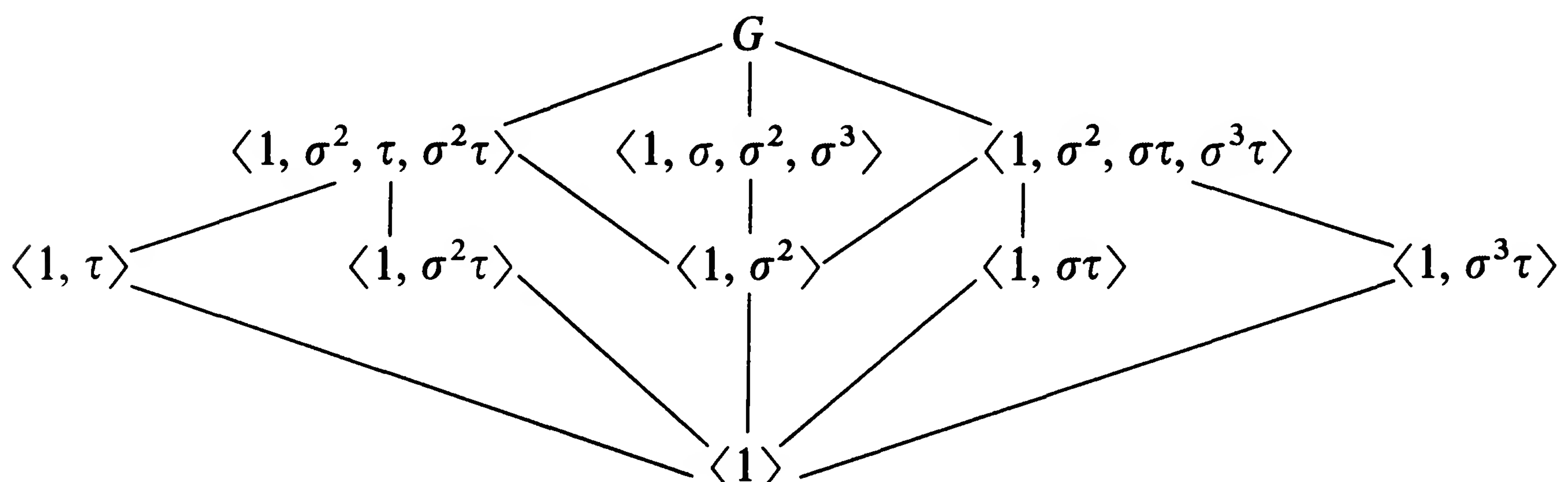
There exists an automorphism τ of K leaving $\mathbf{Q}(\alpha)$ fixed, sending i to $-i$, because K is Galois over $\mathbf{Q}(\alpha)$, of degree 2. Then $\tau^2 = \mathrm{id}$.



By the multiplicativity of degrees in towers, we see that the degrees are as indicated in the diagram. Thus $X^4 - 2$ is irreducible over $\mathbf{Q}(i)$. Also, K is normal over $\mathbf{Q}(i)$. There exists an automorphism σ of K over $\mathbf{Q}(i)$ mapping the root α of $X^4 - 2$ to the root $i\alpha$. Then one verifies at once that $1, \sigma, \sigma^2, \sigma^3$ are distinct and $\sigma^4 = \mathrm{id}$. Thus σ generates a cyclic group of order 4. We denote it by $\langle\sigma\rangle$. Since $\tau \notin \langle\sigma\rangle$ it follows that $G = \langle\sigma, \tau\rangle$ is generated by σ and τ because $\langle\sigma\rangle$ has index 2. Furthermore, one verifies directly that

$$\tau\sigma = \sigma^3\tau,$$

because this relation is true when applied to α and i which generate K over $\mathbf{Q}$. This gives us the structure of G. It is then easy to verify that the lattice of subgroups is as follows:



Example 4. Let k be a field and let $t_1, \dots, t_n$ be algebraically independent over k. Let $K = k(t_1, \dots, t_n)$. The symmetric group G on n letters operates on K by permuting $(t_1, \dots, t_n)$ and its fixed field is the field of symmetric functions, by definition the field of those elements of K fixed under G. Let $s_1, \dots, s_n$ be the elementary symmetric polynomials, and let

$$f(X) = \prod_{i=1}^{n} (X - t_i).$$

Up to a sign, the coefficients of f are $s_1, \dots, s_n$. We let $F = K^G$. We contend that $F = k(s_1, \dots, s_n)$. Indeed,

$$k(s_1, \dots, s_n) \subset F.$$

On the other hand, K is the splitting field of $f(X)$, and its degree over F is $n!$. Its degree over $k(s_1, \dots, s_n)$ is $\leqq n!$ and hence we have equality, $F = k(s_1, \dots, s_n)$.

The polynomial $f(X)$ above is called the general polynomial of degree n. We have just constructed a Galois extension whose Galois group is the symmetric group.

Using the Hilbert irreducibility theorem, one can construct a Galois extension of $\mathbf{Q}$ whose Galois group is the symmetric group. (Cf. Chapter VII, end of §2, and [La 83], Chapter IX.) It is unknown whether given a finite group G, there exists a Galois extension of $\mathbf{Q}$ whose Galois group is G. By specializing parameters, Emmy Noether remarked that one could prove this if one knew that every field E such that

$$\mathbf{Q}(s_1, \dots, s_n) \subset E \subset \mathbf{Q}(t_1, \dots, t_n)$$

is isomorphic to a field generated by n algebraically independent elements. However, matters are not so simple, because Swan proved that the fixed field of a cyclic subgroup of the symmetric group is not necessarily generated by algebraically independent elements over k [Sw 69], [Sw 83].

Example 5. *We shall prove that the complex numbers are algebraically closed.* This will illustrate almost all the theorems we have proved previously.

We use the following properties of the real numbers $\mathbf{R}$: It is an ordered field, every positive element is a square, and every polynomial of odd degree in $\mathbf{R}[X]$ has a root in $\mathbf{R}$. We shall discuss ordered fields in general later, and our arguments apply to any ordered field having the above properties.

Let $i = \sqrt{-1}$ (in other words a root of $X^2 + 1$). Every element in $\mathbf{R}(i)$ has a square root. If $a + bi \in \mathbf{R}(i)$, $a, b \in \mathbf{R}$, then the square root is given by $c + di$, where

$$c^2 = \frac{a + \sqrt{a^2 + b^2}}{2} \quad \text{and} \quad d^2 = \frac{-a + \sqrt{a^2 + b^2}}{2}.$$

Each element on the right of our equalities is positive and hence has a square root in $\mathbf{R}$. It is then trivial to determine the sign of c and d so that $(c + di)^2 = a + bi$.

Since $\mathbf{R}$ has characteristic 0, every finite extension is separable. Every finite extension of $\mathbf{R}(i)$ is contained in an extension K which is finite and Galois over $\mathbf{R}$. We must show that $K = \mathbf{R}(i)$. Let G be the Galois group over $\mathbf{R}$ and let H be a 2-Sylow subgroup of G. Let F be its fixed field. Counting degrees and orders, we find that the degree of F over $\mathbf{R}$ is odd. By the primitive element theorem, there exists an element $\alpha \in F$ such that $F = \mathbf{R}(\alpha)$. Then α is the root of an irreducible polynomial in $\mathbf{R}[X]$ of odd degree. This can happen only if this degree is 1. Hence $G = H$ is a 2-group.

We now see that K is Galois over $\mathbf{R}(i)$. Let G_1 be its Galois group. Since G_1 is a p-group (with $p = 2$), if G_1 is not the trivial group, then G_1 has a subgroup G_2 of index 2. Let F be the fixed field of G_2. Then F is of degree 2 over $\mathbf{R}(i)$; it is a quadratic extension. But we saw that every element of $\mathbf{R}(i)$ has a square root, and hence that $\mathbf{R}(i)$ has no extensions of degree 2. It follows that G_1 is the trivial group and $K = \mathbf{R}(i)$, which is what we wanted.

(The basic ideas of the above proof were already in Gauss. The variation of the ideas which we have selected, making a particularly efficient use of the Sylow group, is due to Artin.)

Example 6. Let $f(X)$ be an irreducible polynomial over the field k, and assume that f is separable. Then the Galois group G of the splitting field is represented as a group of permutations of the n roots, where $n = \deg f$. Whenever one has a criterion for this group to be the full symmetric group S_n, then one can see if it applies to this representation of G. For example, it is an easy exercise (cf. Chapter I, Exercise 38) that for p prime, S_p is generated by $[123 \cdots p]$ and any transposition. We then have the following result.

Let $f(X)$ be an irreducible polynomial with rational coefficients and of degree p prime. If f has precisely two nonreal roots in the complex numbers, then the Galois group of f is S_p.

Proof. The order of G is divisible by p, and hence by Sylow's theorem, G contains an element of order p. Since G is a subgroup of S_p which has order $p!$, it follows that an element of order p can be represented by a p-cycle $[123 \cdots p]$ after a suitable ordering of the roots, because any smaller cycle has order less than p, so relatively prime to p. But the pair of complex conjugate roots shows that complex conjugation induces a transposition in G. Hence the group is all of S_p.

A specific case is easily given. Drawing the graph of

$$f(X) = X^5 - 4X + 2$$

shows that f has exactly three real roots, so exactly two complex conjugate roots. Furthermore f is irreducible over $\mathbf{Q}$ by Eisenstein's criterion, so we can apply the general statement proved above to conclude that the Galois group of f over $\mathbf{Q}$ is S_5. See also Exercise 17 of Chapter IV.

Example 7. The preceding example determines a Galois group by finding some subgroups passing to an extension field of the ground field. There are other possible extensions of **Q** rather than the reals, for instance p-adic fields which will be discussed later in this book. However, instead of passing to an extension field, it is possible to use reduction mod p. For our purposes here, we assume the following statement, which will be proved in Chapter VII, theorem 2.9.

Let $f(X) \in \mathbf{Z}[X]$ be a polynomial with integral coefficients, and leading coefficient 1. *Let p be a prime number. Let $\bar{f}(X) = f(X) \bmod p$ be the polynomial obtained by reducing the coefficients* mod p. *Assume that $\bar{f}$ has no multiple roots in an algebraic closure of $\mathbf{F}_p$. Then there exists a bijection*

$$(\alpha_1, \ldots, \alpha_n) \mapsto (\bar{\alpha}_1, \ldots, \bar{\alpha}_n)$$

of the roots of f onto those of $\bar{f}$, and an embedding of the Galois group of $\bar{f}$ as a subgroup of the Galois group of f, which gives an isomorphism of the action of those groups on the set of roots.

The embedding will be made precise in Chapter VII, but here we just want to use this result to compute Galois groups.

For instance, consider $X^5 - X - 1$ over **Z**. Reducing mod 5 shows that this polynomial is irreducible. Reducing mod 2 gives the irreducible factors

$$(X^2 + X + 1)(X^3 + X^2 + 1) \quad (\text{mod } 2).$$

Hence the Galois group over the rationals contains a 5-cycle and a product of a 2-cycle and a 3-cycle. The third power of the product of the 2-cycle and 3-cycle is a 2-cycle, which is a transposition. Hence the Galois group contains a transposition and the cycle $[123 \cdots p]$, which generate S_p (cf. the exercises of Chapter I on the symmetric group). Thus the Galois group of $X^5 - X - 1$ is S_p.

Example 8. The technique of reducing mod primes to get lots of elements in a Galois group was used by Schur to determine the Galois groups of classical polynomials [Schur 31]. For instance, Schur proves that the Galois group over **Q** of the following polynomials over **Q** is the symmetric group:

(a) $f(X) = \sum_{m=0}^{n} X^m/m!$ (in other words, the truncated exponential series), if n is not divisible by 4. If n is divisible by 4, he gets the alternating group.

(b) Let

$$H_m(X) = (-1)^m e^{X^2/2} \frac{d^m}{dX^m}(e^{-X^2/2})$$

be the m-th Hermite polynomial. Put

$$H_{2n}(X) = K_n^{(0)}(X^2) \quad \text{and} \quad H_{2n+1}(X) = XK_n^{(1)}(X^2).$$

Then the Galois group of $K_n^{(i)}(X)$ over **Q** is the symmetric group S_n for $i = 0$, 1, provided $n > 12$. The remaining cases were settled in [Schulz 37].

Example 9. This example is addressed to those who know something about Riemann surfaces and coverings. Let t be transcendental over the complex numbers $\mathbf{C}$, and let $k = \mathbf{C}(t)$. The values of t in $\mathbf{C}$, or ∞, correspond to the points of the Gauss sphere S, viewed as a Riemann surface. Let $P_1, \dots, P_{n+1}$ be distinct points of S. The finite coverings of $S - \{P_1, \dots, P_{n-1}\}$ are in bijection with certain finite extensions of $\mathbf{C}(t)$, those which are unramified outside $P_1, \dots, P_{n-1}$. Let K be the union of all these extension fields corresponding to such coverings, and let $\pi_1^{(n)}$ be the fundamental group of

$$S - \{P_1, \dots, P_{n+1}\}.$$

Then it is known that $\pi_1^{(n)}$ is a free group on n generators, and has an embedding in the Galois group of K over $\mathbf{C}(t)$, such that the finite subfields of K over $\mathbf{C}(t)$ are in bijection with the subgroups of $\pi_1^{(n)}$ which are of finite index. Given a finite group G generated by n elements $\sigma_1, \dots, \sigma_n$ we can find a surjective homomorphism $\pi_1^{(n)} \to G$ mapping the generators of $\pi_1^{(n)}$ on $\sigma_1, \dots, \sigma_n$. Let H be the kernel. Then H belongs to a subfield K^H of K which is normal over $\mathbf{C}(t)$ and whose Galois group is G. In the language of coverings, H belongs to a finite covering of

$$S - \{P_1, \dots, P_{n+1}\}.$$

Over the field $\mathbf{C}(t)$ one can use analytic techniques to determine the Galois group. The Galois group is the completion of a free group, as proved by Douady [Dou 64]. For extensions to characteristic p, see [Pop 95]. A fundamental problem is to determine the Galois group over $\mathbf{Q}(t)$, which requires much deeper insight into the number theoretic nature of this field. Basic contributions were made by Belyi [Be 80], [Be 83], who also considered the field $\mathbf{Q}(\mu)(t)$, where $\mathbf{Q}(\mu)$ is the field obtained by adjoining all roots of unity to the rationals. Belyi proved that over this latter field, essentially all the classical finite groups occur as Galois groups. See also Conjecture 14.2 below.

For Galois groups over $\mathbf{Q}(t)$, see the survey [Se 88], which contains a bibliography. One method is called the rigidity method, first applied by Shih [Shi 74], which I summarize because it gives examples of various notions defined throughout this book. The problem is to descend extensions of $\mathbf{C}(t)$ with a given Galois group G to extensions of $\mathbf{Q}(t)$ with the same Galois group. If this extension is K over $\mathbf{Q}(t)$, one also wants the extension to be regular over $\mathbf{Q}$ (see the definition in Chapter VIII, §4). To give a sufficient condition, we need some definitions. Let G be a finite group with trivial center. Let C_1, C_2, C_3 be conjugacy classes. Let $P = P(C_1, C_2, C_3)$ be the set of elements

$$(g_1, g_2, g_3) \in C_1 \times C_2 \times C_3$$

such that $g_1 g_2 g_3 = 1$. Let P' be the subset of P consisting of all elements $(g_1, g_2, g_3) \in P$ such that G is generated by g_1, g_2, g_3. We say that the family (C_1, C_2, C_3) is **rigid** if G operates transitively on P', and P' is not empty.

We define a conjugacy class C of G to be **rational** if given $g \in C$ and a positive integer s relatively prime to the order of g, then $g^s \in C$. (Assuming that the reader knows the terminology of characters defined in Chapter XVIII, this condition of rationality is equivalent to the condition that every character χ of G has values in the rational numbers $\mathbf{Q}$.) One then has the following theorem, which is contained in the works of Shih, Fried, Belyi, Matzat and Thompson.

Rigidity theorem. *Let G be a finite group with trivial center, and let C_1, C_2, C_3 be conjugacy classes which are rational, and such that the family (C_1, C_2, C_3) is rigid. Then there exists a Galois extension of $\mathbf{Q}(t)$ with Galois group G (and such that the extension is regular over $\mathbf{Q}$).*

Bibliography

[Be 80] G. BELYI, Galois extensions of the maximal cyclotomic field, *Izv. Akad. Nauk SSR* **43** (1979) pp. 267–276 (= *Math. USSR Izv.* **14** (1980), pp. 247–256

[Be 83] G. BELYI, On extensions of the maximal cyclotomic field having a given classical Galois group, *J. reine angew. Math.* **341** (1983), pp. 147–156

[Dou 64] A. DOUADY, Determination d'un groupe de Galois, *C.R. Acad. Sci.* **258** (1964), pp. 5305–5308

[La 83] S. LANG, *Fundamentals of Diophantine Geometry.* Springer Verlag 1983

[Pop 95] F. POP, Etale Galois covers of affine smooth curves, *Invent. Math.* **120** (1995), pp. 555–578

[Se 88] J.-P.. SERRE, Groupes de Galois sur **Q**, *Séminaire Bourbaki*, 1987–1988 *Astérisque* **161–162**, pp. 73–85

[Shi 74] R.-Y. SHIH, On the construction of Galois extensions of function fields and number fields, *Math. Ann.* **207** (1974), pp. 99–120

[Sw 69] R. SWAN, Invariant rational functions and a problem of Steenrod, *Invent. Math.* **7** (1969), pp. 148–158

[Sw 83] R. SWAN, Noether's problem in Galois theory, *Emmy Noether in Bryn Mawr*, J. D. Sally and B. Srinivasan, eds., Springer Verlag, 1983, pp. 40

§3. ROOTS OF UNITY

Let k be a field. By a **root of unity** (in k) we shall mean an element $\zeta \in k$ such that $\zeta^n = 1$ for some integer $n \geqq 1$. If the characteristic of k is p, then the equation

$$X^{p^m} = 1$$

has only one root, namely 1, and hence there is no p^m-th root of unity except 1.

Let n be an integer > 1 and not divisible by the characteristic. The polynomial

$$X^n - 1$$

is separable because its derivative is $nX^{n-1} \neq 0$, and the only root of the derivative is 0, so there is no common root. Hence in k^a the polynomial $X^n - 1$ has n distinct roots, which are roots of unity. They obviously form a group, and we know that every finite multiplicative group in a field is cyclic (Chapter IV, Theorem 1.9). Thus the group of n-th roots of unity is cyclic. A generator for this group is called a **primitive** n-th root of unity.

If $\boldsymbol{\mu}_n$ denotes the group of all n-th roots of unity in k^a and m, n are relatively prime integers, then

$$\boldsymbol{\mu}_{mn} \approx \boldsymbol{\mu}_m \times \boldsymbol{\mu}_n.$$

This follows because $\boldsymbol{\mu}_m$, $\boldsymbol{\mu}_n$ cannot have any element in common except 1, and because $\boldsymbol{\mu}_m\boldsymbol{\mu}_n$ consequently has mn elements, each of which is an mn-th root of unity. Hence $\boldsymbol{\mu}_m\boldsymbol{\mu}_n = \boldsymbol{\mu}_{mn}$, and the decomposition is that of a direct product.

As a matter of notation, to avoid double indices, especially in the prime power case, we write $\boldsymbol{\mu}[n]$ for $\boldsymbol{\mu}_n$. So if p is a prime, $\boldsymbol{\mu}[p^r]$ is the group of p^r-th roots of unity. Then $\boldsymbol{\mu}[p^\infty]$ denotes the union of all $\boldsymbol{\mu}[p^r]$ for all positive integers r. See the comments in §14.

Let k be any field. Let n be not divisible by the characteristic p. Let $\zeta = \zeta_n$ be a primitive n-th root of unity in k^a. Let σ be an embedding of $k(\zeta)$ in k^a over k. Then

$$(\sigma\zeta)^n = \sigma(\zeta^n) = 1$$

so that $\sigma\zeta$ is an n-th root of unity also. Hence $\sigma\zeta = \zeta^i$ for some integer $i = i(\sigma)$, uniquely determined mod n. It follows that σ maps $k(\zeta)$ into itself, and hence that $k(\zeta)$ is normal over k. If τ is another automorphism of $k(\zeta)$ over k then

$$\sigma\tau\zeta = \zeta^{i(\sigma)i(\tau)}.$$

Since σ and τ are automorphisms, it follows that $i(\sigma)$ and $i(\tau)$ are prime to n (otherwise, $\sigma\zeta$ would have a period smaller than n). In this way we get a homomorphism of the Galois group G of $k(\zeta)$ over k into the multiplicative group $(\mathbf{Z}/n\mathbf{Z})^*$ of integers prime to n, mod n. Our homomorphism is clearly injective since $i(\sigma)$ is uniquely determined by σ mod n, and the effect of σ on $k(\zeta)$ is determined by its effect on ζ. *We conclude that $k(\zeta)$ is abelian over k.*

We know that the order of $(\mathbf{Z}/n\mathbf{Z})^*$ is $\varphi(n)$. Hence the degree $[k(\zeta):k]$ divides $\varphi(n)$.

For a specific field k, the question arises whether the image of $G_{\kappa(\zeta)/\kappa}$ in $(\mathbf{Z}/n\mathbf{Z})^*$ is all of $(\mathbf{Z}/n\mathbf{Z})^*$. Looking at $\kappa = \mathbf{R}$ or $\mathbf{C}$, one sees that this is not always the case. We now give an important example when it is the case.

Theorem 3.1. *Let ζ be a primitive n-th root of unity. Then*

$$[\mathbf{Q}(\zeta):\mathbf{Q}] = \varphi(n),$$

where φ is the Euler function. The map $\sigma \mapsto i(\sigma)$ gives an isomorphism

$$G_{\mathbf{Q}(\zeta)/\mathbf{Q}} \xrightarrow{\approx} (\mathbf{Z}/n\mathbf{Z})^*.$$

Proof. Let $f(X)$ be the irreducible polynomial of ζ over $\mathbf{Q}$. Then $f(X)$ divides $X^n - 1$, say $X^n - 1 = f(X)h(X)$, where both f, h have leading coefficient 1. By the Gauss lemma, it follows that f, h have integral coefficients. We shall now prove that if p is a prime number not dividing n, then ζ^p is also a root of f. Since ζ^p is also a primitive n-th root of unity, and since any primitive n-th root of unity can be obtained by raising ζ to a succession of prime powers, with primes not dividing n, this will imply that all the primitive n-th roots of unity are roots of f, which must therefore have degree $\geqq \varphi(n)$, and hence precisely $\varphi(n)$.

Suppose ζ^p is not a root of f. Then ζ^p is a root of h, and ζ itself is a root of $h(X^p)$. Hence $f(X)$ divides $h(X^p)$, and we can write

$$h(X^p) = f(X)g(X).$$

Since f has integral coefficients and leading coefficient 1, we see that g has integral coefficients. Since $a^p \equiv a \pmod{p}$ for any integer a, we conclude that

$$h(X^p) \equiv h(X)^p \pmod{p},$$

and hence

$$h(X)^p \equiv f(X)g(X) \pmod{p}.$$

In particular, if we denote by $\bar{f}$ and $\bar{h}$ the polynomials in $\mathbf{Z}/p\mathbf{Z}$ obtained by reducing f and h respectively mod p, we see that $\bar{f}$ and $\bar{h}$ are not relatively prime, i.e. have a factor in common. But $X^n - \bar{1} = \bar{f}(X)\bar{h}(X)$, and hence $X^n - \bar{1}$ has multiple roots. This is impossible, as one sees by taking the derivative, and our theorem is proved.

Corollary 3.2. *If n, m are relative prime integers $\geqq 1$, then*

$$\mathbf{Q}(\zeta_n) \cap \mathbf{Q}(\zeta_m) = \mathbf{Q}.$$

Proof. We note that ζ_n and ζ_m are both contained in $\mathbf{Q}(\zeta_{mn})$ since ζ_{mn}^n is a primitive m-th root of unity. Furthermore, $\zeta_m\zeta_n$ is a primitive mn-th root of unity. Hence

$$\mathbf{Q}(\zeta_n)\mathbf{Q}(\zeta_m) = \mathbf{Q}(\zeta_{mn}).$$

Our assertion follows from the multiplicativity $\varphi(mn) = \varphi(m)\varphi(n)$.

Suppose that n is a prime number p (having nothing to do with the characteristic). Then

$$X^p - 1 = (X - 1)(X^{p-1} + \cdots + 1).$$

Any primitive p-th root of unity is a root of the second factor on the right of this equation. Since there are exactly $p - 1$ primitive p-th roots of unity, we conclude that these roots are precisely the roots of

$$X^{p-1} + \cdots + 1.$$

We saw in Chapter IV, §3 that this polynomial could be transformed into an Eisenstein polynomial over the rationals. This gives another proof that $[\mathbf{Q}(\zeta_p) : \mathbf{Q}] = p - 1$.

We investigate more closely the factorization of $X^n - 1$, and suppose that we are in characteristic 0 for simplicity.

We have

$$X^n - 1 = \prod_{\zeta} (X - \zeta),$$

where the product is taken over all n-th roots of unity. Collect together all terms belonging to roots of unity having the same period. Let

$$\Phi_d(X) = \prod_{\text{period } \zeta = d} (X - \zeta)$$

Then

$$X^n - 1 = \prod_{d|n} \Phi_d(X).$$

We see that $\Phi_1(X) = X - 1$, and that

$$\Phi_n(X) = \frac{X^n - 1}{\prod_{\substack{d|n \\ d<n}} \Phi_d(X)}.$$

From this we can compute $\Phi(X)$ recursively, and we see that $\Phi_n(X)$ is a polynomial in $\mathbf{Q}[X]$ because we divide recursively by polynomials having coefficients in $\mathbf{Q}$. All our polynomials have leading coefficient 1, so that in fact $\Phi_n(X)$ has *integer coefficients* by Theorem 1.1 of Chapter IV. Thus our construction is essentially universal and would hold over any field (whose characteristic does not divide n).

We call $\Phi_n(X)$ the n-th **cyclotomic polynomial**.

The roots of Φ_n are precisely the primitive n-th roots of unity, and hence

$$\deg \Phi_n = \varphi(n).$$

From Theorem 3.1 we conclude that Φ_n is irreducible over $\mathbf{Q}$, *and hence*

$$\Phi_n(X) = \operatorname{Irr}(\zeta_n, \mathbf{Q}, X).$$

We leave the proofs of the following recursion formulas as exercises:

1. If p is a prime number, then

$$\Phi_p(X) = X^{p-1} + X^{p-2} + \cdots + 1,$$

and for an integer $r \geqq 1$,

$$\Phi_{p^r}(X) = \Phi_p(X^{p^{r-1}}).$$

2. Let $n = p_1^{r_1} \cdots p_s^{r_s}$ be a positive integer with its prime factorization. Then

$$\Phi_n(X) = \Phi_{p_1 \cdots p_s}(X^{p_1^{r_1-1} \cdots p_s^{r_s-1}}).$$

3. If n is odd > 1, then $\Phi_{2n}(X) = \Phi_n(-X)$.
4. If p is a prime number, not dividing n, then

$$\Phi_{pn}(X) = \frac{\Phi_n(X^p)}{\Phi_n(X)}.$$

On the other hand, if $p \mid n$, then $\Phi_{pn}(X) = \Phi_n(X^p)$.

5. We have

$$\Phi_n(X) = \prod_{d \mid n} (X^{n/d} - 1)^{\mu(d)}.$$

As usual, μ is the Möbius function:

$$\mu(n) = \begin{cases} 0 & \text{if } n \text{ is divisible by } p^2 \text{ for some prime } p, \\ (-1)^r & \text{if } n = p_1 \cdots p_r \text{ is a product of distinct primes,} \\ 1 & \text{if } n = 1. \end{cases}$$

As an exercise, show that

$$\sum_{d \mid n} \mu(d) = \begin{cases} 1 & \text{if } n = 1, \\ 0 & \text{if } n > 1. \end{cases}$$

Example. In light of Exercise 21 of Chapter V, we note that the association $n \mapsto \Phi_n(X)$ can be viewed as a function from the positive integers into the multiplicative group of non-zero rational functions. The multiplication formula $X^n - 1 = \prod \Phi_d(X)$ can therefore be inverted by the general formalism of convolutions. Computations of a number of cyclotomic polynomials show that for low values of n, they have coefficients equal to 0 or ± 1. However, I am indebted to Keith Conrad for bringing to my attention an extensive literature on the subject, starting with Bang in 1895. I include only the first and last items:

A. S. BANG, Om Ligningen $\Phi_m(X) = 0$, *Nyt Tidsskrift for Matematik* **(B) 6** (1895), pp. 6–12

H. L. MONTGOMERY and R. C. VAUGHN, The order of magnitude of the m-th coefficients of cyclotomic polynomials, *Glasgow Math. J.* **27** (1985), pp. 143–159

In particular, if $\Phi_n(X) = \Sigma a_{nj}X^j$, define $L(j) = \log \max_n |a_{nj}|$. Then Montgomery and Vaughn prove that

$$\frac{j^{1/2}}{(\log j)^{1/4}} \ll L(j) \ll \frac{j^{1/2}}{(\log j)^{1/4}}$$

where the sign $\ll$ means that the left-hand side is at most a positive constant times the right-hand side for $j \to \infty$. Bang also points out that $\Phi_{105}(X)$ is a cyclotomic polynomial of smallest degree having coefficients $\neq 0$ or ± 1: the coefficient of X^7 and X^{41} is -2 (all others are 0 or ± 1).

If ζ is an n-th root of unity and $\zeta \neq 1$, then

$$\frac{1 - \zeta^n}{1 - \zeta} = 1 + \zeta + \cdots + \zeta^{n-1} = 0.$$

This is trivial, but useful.

Let $\mathbf{F}_q$ be the finite field with q elements, q equal to a power of the odd prime number p. Then $\mathbf{F}_q^*$ has $q - 1$ elements and is a cyclic group. Hence we have the index

$$(\mathbf{F}_q^* : \mathbf{F}_q^{*2}) = 2.$$

If ν is a non-zero integer not divisible by p, let

$$\left(\frac{\nu}{p}\right) = \begin{cases} 1 & \text{if } \nu \equiv x^2 \pmod{p} \text{ for some } x, \\ -1 & \text{if } \nu \not\equiv x^2 \pmod{p} \text{ for all } x. \end{cases}$$

This is known as the **quadratic symbol**, and depends only on the residue class of ν mod p.

From our preceding remark, we see that there are as many quadratic residues as there are non-residues mod p.

Theorem 3.3. *Let ζ be a primitive p-th root of unity, and let*

$$S = \sum_{\nu} \left(\frac{\nu}{p}\right)\zeta^{\nu},$$

the sum being taken over non-zero residue classes mod p. Then

$$S^2 = \left(\frac{-1}{p}\right)p.$$

Every quadratic extension of $\mathbf{Q}$ is contained in a cyclotomic extension.

Proof. The last statement follows at once from the explicit expression of $\pm p$ as a square in $\mathbf{Q}(\zeta)$, because the square root of an integer is contained in the

field obtained by adjoining the square root of the prime factors in its factorization, and also $\sqrt{-1}$. Furthermore, for the prime 2, we have $(1 + i)^2 = 2i$. We now prove our assertion concerning S^2. We have

$$S^2 = \sum_{\nu,\mu} \left(\frac{\nu}{p}\right)\left(\frac{\mu}{p}\right)\zeta^{\nu+\mu} = \sum_{\nu,\mu} \left(\frac{\nu\mu}{p}\right)\zeta^{\nu+\mu}.$$

As ν ranges over non-zero residue classes, so does $\nu\mu$ for any fixed μ, and hence replacing ν by $\nu\mu$ yields

$$\begin{aligned} S^2 &= \sum_{\nu,\mu} \left(\frac{\nu\mu^2}{p}\right)\zeta^{\mu(\nu+1)} = \sum_{\nu,\mu} \left(\frac{\nu}{p}\right)\zeta^{\mu(\nu+1)} \\ &= \sum_{\mu} \left(\frac{-1}{p}\right)\zeta^0 + \sum_{\nu \neq -1} \left(\frac{\nu}{p}\right) \sum_{\mu} \zeta^{\mu(\nu+1)}. \end{aligned}$$

But $1 + \zeta + \cdots + \zeta^{p-1} = 0$, and the sum on the right over μ consequently yields -1. Hence

$$\begin{aligned} S^2 &= \left(\frac{-1}{p}\right)(p-1) + (-1)\sum_{\nu \neq -1} \left(\frac{\nu}{p}\right) \\ &= p\left(\frac{-1}{p}\right) - \sum_{\nu} \left(\frac{\nu}{p}\right) \\ &= p\left(\frac{-1}{p}\right), \end{aligned}$$

as desired.

We see that $\mathbf{Q}(\sqrt{p})$ is contained in $\mathbf{Q}(\zeta, \sqrt{-1})$ or $\mathbf{Q}(\zeta)$, depending on the sign of the quadratic symbol with -1. An extension of a field is said to be **cyclotomic** if it is contained in a field obtained by adjoining roots of unity. We have shown above that quadratic extensions of $\mathbf{Q}$ are cyclotomic. A theorem of Kronecker asserts that every abelian extension of $\mathbf{Q}$ is cyclotomic, but the proof needs techniques which cannot be covered in this book.

§4. LINEAR INDEPENDENCE OF CHARACTERS

Let G be a monoid and K a field. By a **character** of G in K (in this chapter), we shall mean a homomorphism

$$\chi: G \to K^*$$

of G into the multiplicative group of K. The **trivial character** is the homo-

morphism taking the constant value 1. Functions $f_i: G \to K$ are called **linearly independent** over K if whenever we have a relation

$$a_1 f_1 + \cdots + a_n f_n = 0$$

with $a_i \in K$, then all $a_i = 0$.

Examples. Characters will occur in various contexts in this book. First, the various conjugate embeddings of an extension field in an algebraic closure can be viewed as characters. These are the characters which most concern us in this chapter. Second, we shall meet characters in Chapter XVIII, when we shall extend the next theorem to a more general kind of character in connection with group representations.

Next, one meets characters in analysis. For instance, given an integer m, the function $f: \mathbf{R}/\mathbf{Z} \to \mathbf{C}^*$ such that $f(x) = e^{2\pi i m x}$ is a character on $\mathbf{R}/\mathbf{Z}$. It can be shown that all continuous homomorphisms of $\mathbf{R}/\mathbf{Z}$ into $\mathbf{C}^*$ are of this type. Similarly, given a real number y, the function $x \mapsto e^{2\pi i x y}$ is a continuous character on $\mathbf{R}$, and it is shown in Fourier analysis that all continuous characters of absolute value 1 on $\mathbf{R}$ are of this type.

Further, let X be a compact space and let R be the ring of continuous complex-valued functions on X. Let R^* be the group of units of R. Then given $x \in X$ the evaluation map $f \mapsto f(x)$ is a character of R^* into $\mathbf{C}^*$. (Actually, this evaluation map is a ring homomorphism of R onto $\mathbf{C}$.)

Artin found a neat way of expressing a linear independence property which covers all these cases, as well as others, in the following theorem [Ar 44].

Theorem 4.1. (Artin). *Let G be a monoid and K a field. Let $\chi_1, \ldots, \chi_n$ be distinct characters of G in K. Then they are linearly independent over K.*

Proof. One character is obviously linearly independent. Suppose that we have a relation

$$a_1\chi_1 + \cdots + a_n\chi_n = 0$$

with $a_i \in K$, not all 0. Take such a relation with n as small as possible. Then $n \geqq 2$, and no a_i is equal to 0. Since χ_1, χ_2 are distinct, there exists $z \in G$ such that $\chi_1(z) \neq \chi_2(z)$. For all $x \in G$ we have

$$a_1\chi_1(xz) + \cdots + a_n\chi_n(xz) = 0,$$

and since χ_i is a character,

$$a_1\chi_1(z)\chi_1 + \cdots + a_n\chi_n(z)\chi_n = 0.$$

Divide by $\chi_1(z)$ and subtract from our first relation. The term $a_1\chi_1$ cancels, and we get a relation

$$\left(a_2\frac{\chi_2(z)}{\chi_1(z)} - a_2\right)\chi_2 + \cdots = 0.$$

The first coefficient is not 0, and this is a relation of smaller length than our first relation, contradiction.

As an application of Artin's theorem, one can consider the case when K is a finite normal extension of a field k, and when the characters are distinct automorphisms $\sigma_1, \ldots, \sigma_n$ of K over k, viewed as homomorphisms of K^* into K^*. This special case had already been considered by Dedekind, who, however, expressed the theorem in a somewhat different way, considering the determinant constructed from $\sigma_i \omega_j$ where ω_j is a suitable set of elements of K, and proving in a more complicated way the fact that this determinant is not 0. The formulation given above and its particularly elegant proof are due to Artin.

As another application, we have:

Corollary 4.2. *Let $\alpha_1, \ldots, \alpha_n$ be distinct non-zero elements of a field K. If $a_1, \ldots, a_n$ are elements of K such that for all integers $\nu \geqq 0$ we have*

$$a_1\alpha_1^\nu + \cdots + a_n\alpha_n^\nu = 0$$

then $a_i = 0$ for all i.

Proof. We apply the theorem to the distinct homomorphisms

$$\nu \mapsto \alpha_i^\nu$$

of $\mathbf{Z}_{\geqq 0}$ into K^*.

Another interesting application will be given as an exercise (relative invariants).

§5. THE NORM AND TRACE

Let E be a finite extension of k. Let $[E:k]_s = r$, and let

$$p^\mu = [E:k]_i$$

if the characteristic is $p > 0$, and 1 otherwise. Let $\sigma_1, \ldots, \sigma_r$ be the distinct embeddings of E in an algebraic closure k^a of k. If α is an element of E, we define its **norm** from E to k to be

$$N_{E/k}(\alpha) = N_k^E(\alpha) = \prod_{\nu=1}^{r} \sigma_\nu \alpha^{p^\mu} = \left(\prod_{\nu=1}^{r} \sigma_\nu \alpha\right)^{[E:k]_i}.$$

Similarly, we define the **trace**

$$\mathrm{Tr}_{E/k}(\alpha) = \mathrm{Tr}_k^E(\alpha) = [E:k]_i \sum_{\nu=1}^{r} \sigma_\nu \alpha.$$

The trace is equal to 0 if $[E:k]_i > 1$, in other words, if E/k is not separable.

Thus if E is separable over k, we have

$$N_k^E(\alpha) = \prod_\sigma \sigma\alpha$$

where the product is taken over the distinct embeddings of E in k^a over k.

Similarly, if E/k is separable, then

$$\mathrm{Tr}_k^E(\alpha) = \sum_\sigma \sigma\alpha.$$

Theorem 5.1. *Let E/k be a finite extension. Then the norm N_k^E is a multiplicative homomorphism of E^* into k^* and the trace is an additive homomorphism of E into k. If $E \supset F \supset k$ is a tower of fields, then the two maps are transitive, in other words,*

$$N_k^E = N_k^F \circ N_F^E \quad \textit{and} \quad \mathrm{Tr}_k^E = \mathrm{Tr}_k^F \circ \mathrm{Tr}_F^E.$$

If $E = k(\alpha)$, and $f(X) = \mathrm{Irr}(\alpha, k, X) = X^n + a_{n-1}X^{n-1} + \cdots + a_0$, then

$$N_k^{k(\alpha)}(\alpha) = (-1)^n a_0 \quad \textit{and} \quad \mathrm{Tr}_k^{k(\alpha)}(\alpha) = -a_{n-1}.$$

Proof. For the first assertion, we note that α^{p^μ} is separable over k if $p^\mu = [E:k]_i$. On the other hand, the product

$$\prod_{\nu=1}^r \sigma_\nu \alpha^{p^\mu}$$

is left fixed under any isomorphism into k^a because applying such an isomorphism simply permutes the factors. Hence this product must lie in k since α^{p^μ} is separable over k. A similar reasoning applies to the trace.

For the second assertion, let $\{\tau_j\}$ be the family of distinct embeddings of F into k^a over k. Extend each τ_j to an automorphism of k^a, and denote this extension by τ_j also. Let $\{\sigma_i\}$ be the family of embeddings of E in k^a over F. (Without loss of generality, we may assume that $E \subset k^a$.) If σ is an embedding of E over k in k^a, then for some j, $\tau_j^{-1}\sigma$ leaves F fixed, and hence $\tau_j^{-1}\sigma = \sigma_i$ for some i. Hence $\sigma = \tau_j\sigma_i$ and consequently the family $\{\tau_j\sigma_i\}$ gives all distinct embeddings of E into k^a over k. Since the inseparability degree is multiplicative in towers, our assertion concerning the transitivity of the norm and trace is obvious, because we have already shown that N_F^E maps E into F, and similarly for the trace.

Suppose now that $E = k(\alpha)$. We have

$$f(X) = ((X - \alpha_1)\cdots(X - \alpha_r))^{[E:k]_i}$$

if $\alpha_1, \ldots, \alpha_r$ are the distinct roots of f. Looking at the constant term of f gives us the expression for the norm, and looking at the next to highest term gives us the expression for the trace.

We observe that the trace is a k-linear map of E into k, namely

$$\mathrm{Tr}_k^E(c\alpha) = c\,\mathrm{Tr}_k^E(\alpha)$$

for all $\alpha \in E$ and $c \in k$. This is clear since c is fixed under every embedding of E over k. Thus the trace is a k-linear functional of E into k. For simplicity, we write $\mathrm{Tr} = \mathrm{Tr}_k^E$.

Theorem 5.2. *Let E be a finite separable extension of k. Then* $\mathrm{Tr}: E \to k$ *is a non-zero functional. The map*

$$(x, y) \mapsto \mathrm{Tr}(xy)$$

of $E \times E \to k$ is bilinear, and identifies E with its dual space.

Proof. That Tr is non-zero follows from the theorem on linear independence of characters. For each $x \in E$, the map

$$\mathrm{Tr}_x : E \to k$$

such that $\mathrm{Tr}_x(y) = \mathrm{Tr}(xy)$ is obviously a k-linear map, and the map

$$x \mapsto \mathrm{Tr}_x$$

is a k-homomorphism of E into its dual space $E^\vee$. (We don't write E^* for the dual space because we use the star to denote the multiplicative group of E.) If Tr_x is the zero map, then $\mathrm{Tr}(xE) = 0$. If $x \neq 0$ then $xE = E$. Hence the kernel of $x \mapsto \mathrm{Tr}_x$ is 0. Hence we get an injective homomorphism of E into the dual space $\hat{E}$. Since these spaces have the same finite dimension, it follows that we get an isomorphism. This proves our theorem.

Corollary 5.3. *Let $\omega_1, \dots, \omega_n$ be a basis of E over k. Then there exists a basis $\omega'_1, \dots, \omega'_n$ of E over k such that* $\mathrm{Tr}(\omega_i \omega'_j) = \delta_{ij}$.

Proof. The basis $\omega'_1, \dots, \omega'_n$ is none other than the dual basis which we defined when we considered the dual space of an arbitrary vector space.

Corollary 5.4. *Let E be a finite separable extension of k, and let $\sigma_1, \dots, \sigma_n$ be the distinct set of embeddings of E into k^a over k. Let $w_1, \dots, w_n$ be elements of E. Then the vectors*

$$\xi_1 = (\sigma_1 w_1, \dots, \sigma_1 w_n),$$

$$\dots$$

$$\xi_n = (\sigma_n w_1, \dots, \sigma_n w_n)$$

are linearly independent over E if $w_1, \dots, w_n$ form a basis of E over k.

Proof. Assume that $w_1, \dots, w_n$ form a basis of E/k. Let $\alpha_1, \dots, \alpha_n$ be elements of E such that

$$\alpha_1 \xi_1 + \cdots + \alpha_n \xi_n = 0.$$

Then we see that

$$\alpha_1 \sigma_1 + \cdots + \alpha_n \sigma_n$$

applied to each one of $w_1, \dots, w_n$ gives the value 0. But $\sigma_1, \dots, \sigma_n$ are linearly independent as characters of the multiplicative group E^* into k^{a*}. It follows that $\alpha_i = 0$ for $i = 1, \dots, n$, and our vectors are linearly independent.

Remark. In characteristic 0, one sees much more trivially that the trace is not identically 0. Indeed, if $c \in k$ and $c \neq 0$, then $\mathrm{Tr}(c) = nc$ where $n = [E:k]$, and $n \neq 0$. This argument also holds in characteristic p when n is prime to p.

Proposition 5.5. *Let $E = k(\alpha)$ be a separable extension. Let*

$$f(X) = \mathrm{Irr}(\alpha, k, X),$$

and let $f'(X)$ be its derivative. Let

$$\frac{f(X)}{(X - \alpha)} = \beta_0 + \beta_1 X + \cdots + \beta_{n-1}X^{n-1}$$

with $\beta_i \in E$. Then the dual basis of $1, \alpha, \dots, \alpha^{n-1}$ is

$$\frac{\beta_0}{f'(\alpha)}, \dots, \frac{\beta_{n-1}}{f'(\alpha)}.$$

Proof. Let $\alpha_1, \dots, \alpha_n$ be the distinct roots of f. Then

$$\sum_{i=1}^{n} \frac{f(X)}{(X - \alpha_i)} \frac{\alpha_i^r}{f'(\alpha_i)} = X^r \qquad \text{for} \quad 0 \leqq r \leqq n - 1.$$

To see this, let $g(X)$ be the difference of the left- and right-hand side of this equality. Then g has degree $\leqq n - 1$, and has n roots $\alpha_1, \dots, \alpha_n$. Hence g is identically zero.

The polynomials

$$\frac{f(X)}{(X - \alpha_i)} \frac{\alpha_i^r}{f'(\alpha_i)}$$

are all conjugate to each other. If we define the trace of a polynomial with coefficients in E to be the polynomial obtained by applying the trace to the coefficients, then

$$\mathrm{Tr}\left[\frac{f(X)}{(X - \alpha)} \frac{\alpha^r}{f'(\alpha)}\right] = X^r.$$

Looking at the coefficients of each power of X in this equation, we see that

$$\mathrm{Tr}\left(\alpha^i \frac{\beta_j}{f'(\alpha)}\right) = \delta_{ij},$$

thereby proving our proposition.

Finally we establish a connection with determinants, whose basic properties we now assume. Let E be a finite extension of k, which we view as a finite dimensional vector space over k. For each $\alpha \in E$ we have the k-linear map

multiplication by α,

$$m_\alpha: E \to E \quad \text{such that} \quad m_\alpha(x) = \alpha x.$$

Then we have the determinant $\det(m_\alpha)$, which can be computed as the determinant of the matrix M_α representing m_α with respect to a basis. Similarly we have the trace $\mathrm{Tr}(m_\alpha)$, which is the sum of the diagonal elements of the matrix M_α.

Proposition 5.6. *Let E be a finite extension of k and let $\alpha \in E$. Then*

$$\det(m_\alpha) = N_{E/k}(\alpha) \quad \textit{and} \quad \mathrm{Tr}(m_\alpha) = \mathrm{Tr}_{E/k}(\alpha).$$

Proof. Let $F = k(\alpha)$. If $[F : k] = d$, then $1, \alpha, \ldots, \alpha^{d-1}$ is a basis for F over k. Let $\{w_1, \ldots, w_r\}$ be a basis for E over F. Then $\{\alpha^i w_j\}$ $(i = 0, \ldots, d-1; j = 1, \ldots, r)$ is a basis for E over k. Let

$$f(X) = X^d + a_{d-1}X^{d-1} + \ldots + a_0$$

be the irreducible polynomial of α over k. Then $N_{F/k}(\alpha) = (-1)^d a_0$, and by the transitivity of the norm, we have

$$N_{E/k}(\alpha) = N_{F/k}(\alpha)^r.$$

The reader can verify directly on the above basis that $N_{F/k}(\alpha)^r$ is the determinant of m_α on F, and then that $N_{F/k}(\alpha)^d$ is the determinant of m_α on E, thus concluding the proof for the determinant. The trace is handled exactly in the same way, except that $\mathrm{Tr}_{E/k}(\alpha) = r \cdot \mathrm{Tr}_{F/k}(\alpha)$. The trace of the matrix for m_α on F is equal to $-a_{d-1}$. From this the statement identifying the two traces is immediate, as it was for the norm.

§6. CYCLIC EXTENSIONS

We recall that a finite extension is said to be cyclic if it is Galois and its Galois group is cyclic. The determination of cyclic extensions when enough roots of unity are in the ground field is based on the following fact.

Theorem 6.1. (Hilbert's Theorem 90). *Let K/k be cyclic of degree n with Galois group G. Let σ be a generator of G. Let $\beta \in K$. The norm $N_k^K(\beta) = N(\beta)$ is equal to 1 if and only if there exists an element $\alpha \neq 0$ in K such that $\beta = \alpha/\sigma\alpha$.*

Proof. Assume such an element α exists. Taking the norm of β we get $N(\alpha)/N(\sigma\alpha)$. But the norm is the product over all automorphisms in G. Inserting σ just permutes these automorphisms. Hence the norm is equal to 1.

It will be convenient to use an exponential notation as follows. If $\tau, \tau' \in G$ and $\xi \in K$ we write

$$\xi^{\tau+\tau'} = \xi^\tau \xi^{\tau'}.$$

By Artin's theorem on characters, the map given by

$$\mathrm{id} + \beta\sigma + \beta^{1+\sigma}\sigma^2 + \cdots + \beta^{1+\sigma+\cdots+\sigma^{n-2}}\sigma^{n-1}$$

on K is not identically zero. Hence there exists $\theta \in K$ such that the element

$$\alpha = \theta + \beta\theta^{\sigma} + \beta^{1+\sigma}\theta^{\sigma^2} + \cdots + \beta^{1+\sigma+\cdots+\sigma^{n-2}}\theta^{\sigma^{n-1}}$$

is not equal to 0. It is then clear that $\beta\alpha^{\sigma} = \alpha$ using the fact that $N(\beta) = 1$, and hence that when we apply σ to the last term in the sum, we obtain θ. We divide by α^{σ} to conclude the proof.

Theorem 6.2. *Let k be a field, n an integer > 0 prime to the characteristic of k, and assume that there is a primitive n-th root of unity in k.*

(i) *Let K be a cyclic extension of degree n. Then there exists $\alpha \in K$ such that $K = k(\alpha)$, and α satisfies an equation $X^n - a = 0$ for some $a \in k$.*

(ii) *Conversely, let $a \in k$. Let α be a root of $X^n - a$. Then $k(\alpha)$ is cyclic over k, of degree d, $d \mid n$, and α^d is an element of k.*

Proof. Let ζ be a primitive n-th root of unity in k, and let K/k be cyclic with group G. Let σ be a generator of G. We have $N(\zeta^{-1}) = (\zeta^{-1})^n = 1$. By Hilbert's theorem 90, there exists $\alpha \in K$ such that $\sigma\alpha = \zeta\alpha$. Since ζ is in k, we have $\sigma^i\alpha = \zeta^i\alpha$ for $i = 1, \ldots, n$. Hence the elements $\zeta^i\alpha$ are n distinct conjugates of α over k, whence $[k(\alpha) : k]$ is at least equal to n. Since $[K : k] = n$, it follows that $K = k(\alpha)$. Furthermore,

$$\sigma(\alpha^n) = \sigma(\alpha)^n = (\zeta\alpha)^n = \alpha^n.$$

Hence α^n is fixed under σ, hence is fixed under each power of σ, hence is fixed under G. Therefore α^n is an element of k, and we let $a = \alpha^n$. This proves the first part of the theorem.

Conversely, let $a \in k$. Let α be a root of $X^n - a$. Then $\alpha\zeta^i$ is also a root for each $i = 1, \ldots, n$, and hence all roots lie in $k(\alpha)$ which is therefore normal over k. All the roots are distinct so $k(\alpha)$ is Galois over k. Let G be the Galois group.

If σ is an automorphism of $k(\alpha)/k$ then $\sigma\alpha$ is also a root of $X^n - a$. Hence $\sigma\alpha = \omega_\sigma\alpha$ where ω_σ is an n-th root of unity, not necessarily primitive. The map $\sigma \mapsto \omega_\sigma$ is obviously a homomorphism of G into the group of n-th roots of unity, and is injective. Since a subgroup of a cyclic group is cyclic, we conclude that G is cyclic, of order d, and $d \mid n$. The image of G is a cyclic group of order d. If σ is a generator of G, then ω_σ is a primitive dth root of unity. Now we get

$$\sigma(\alpha^d) = (\sigma\alpha)^d = (\omega_\sigma\alpha)^d = \alpha^d.$$

Hence α^d is fixed under σ, and therefore fixed under G. It is an element of k, and our theorem is proved.

We now pass to the analogue of Hilbert's theorem 90 in characteristic p for cyclic extensions of degree p.

Theorem 6.3. (Hilbert's Theorem 90, Additive Form). *Let k be a field and K/k a cyclic extension of degree n with group G. Let σ be a generator of G. Let $\beta \in K$. The trace $\mathrm{Tr}_k^K(\beta)$ is equal to 0 if and only if there exists an element $\alpha \in K$ such that $\beta = \alpha - \sigma\alpha$.*

Proof. If such an element α exists, then we see that the trace is 0 because the trace is equal to the sum taken over all elements of G, and applying σ permutes these elements.

Conversely, assume $\mathrm{Tr}(\beta) = 0$. There exists an element $\theta \in K$ such that $\mathrm{Tr}(\theta) \neq 0$. Let

$$\alpha = \frac{1}{\mathrm{Tr}(\theta)} [\beta\theta^{\sigma} + (\beta + \sigma\beta)\theta^{\sigma^2} + \cdots + (\beta + \sigma\beta + \cdots + \sigma^{n-2}\beta)\theta^{\sigma^{n-1}}].$$

From this it follows at once that $\beta = \alpha - \sigma\alpha$.

Theorem 6.4. (Artin-Schreier) *Let k be a field of characteristic p.*

(i) *Let K be a cyclic extension of k of degree p. Then there exists $\alpha \in K$ such that $K = k(\alpha)$ and α satisfies an equation $X^p - X - a = 0$ with some $a \in k$.*

(ii) *Conversely, given $a \in k$, the polynomial $f(X) = X^p - X - a$ either has one root in k, in which case all its roots are in k, or it is irreducible. In this latter case, if α is a root then $k(\alpha)$ is cyclic of degree p over k.*

Proof. Let K/k be cyclic of degree p. Then $\mathrm{Tr}_k^K(-1) = 0$ (it is just the sum of -1 with itself p times). Let σ be a generator of the Galois group. By the additive form of Hilbert's theorem 90, there exists $\alpha \in K$ such that $\sigma\alpha - \alpha = 1$, or in other words, $\sigma\alpha = \alpha + 1$. Hence $\sigma^i\alpha = \alpha + i$ for all integers $i = 1, \ldots, p$ and α has p distinct conjugates. Hence $[k(\alpha) : k] \geqq p$. It follows that $K = k(\alpha)$. We note that

$$\sigma(\alpha^p - \alpha) = \sigma(\alpha)^p - \sigma(\alpha) = (\alpha + 1)^p - (\alpha + 1) = \alpha^p - \alpha.$$

Hence $\alpha^p - \alpha$ is fixed under σ, hence it is fixed under the powers of σ, and therefore under G. It lies in the fixed field k. If we let $a = \alpha^p - \alpha$ we see that our first assertion is proved.

Conversely, let $a \in k$. If α is a root of $X^p - X - a$ then $\alpha + i$ is also a root for $i = 1, \ldots, p$. Thus $f(X)$ has p distinct roots. If one root lies in k then all roots lie in k. Assume that no root lies in k. We contend that the

polynomial is irreducible. Suppose that

$$f(X) = g(X)h(X)$$

with $g, h \in k[X]$ and $1 \leqq \deg g < p$. Since

$$f(X) = \prod_{i=1}^{p} (X - \alpha - i)$$

we see that $g(X)$ is a product over certain integers i. Let $d = \deg g$. The coefficient of X^{d-1} in g is a sum of terms $-(\alpha + i)$ taken over precisely d integers i. Hence it is equal to $-d\alpha + j$ for some integer j. But $d \neq 0$ in k, and hence α lies in k, because the coefficients of g lie in k, contradiction. We know therefore that $f(X)$ is irreducible. All roots lie in $k(\alpha)$, which is therefore normal over k. Since $f(X)$ has no multiple roots, it follows that $k(\alpha)$ is Galois over k. There exists an automorphism σ of $k(\alpha)$ over k such that $\sigma\alpha = \alpha + 1$ (because $\alpha + 1$ is also a root). Hence the powers σ^i of σ give $\sigma^i\alpha = \alpha + i$ for $i = 1, \ldots, p$ and are distinct. Hence the Galois group consists of these powers and is cyclic, thereby proving the theorem.

For cyclic extensions of degree p^r, see the exercises on Witt vectors and the bibliography at the end of §8.

§7. SOLVABLE AND RADICAL EXTENSIONS

A finite extension E/k (which we shall assume separable for convenience) is said to be **solvable** if the Galois group of the smallest Galois extension K of k containing E is a solvable group. This is equivalent to saying that there exists a solvable Galois extension L of k such that $k \subset E \subset L$. Indeed, we have $k \subset E \subset K \subset L$ and $G(K/k)$ is a homomorphic image of $G(L/k)$.

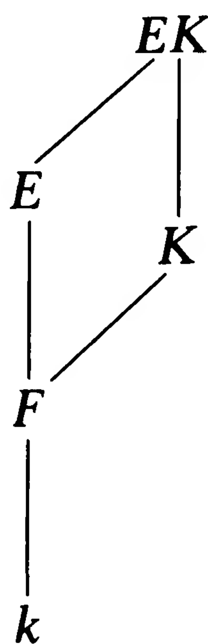
Proposition 7.1. *Solvable extensions form a distinguished class of extensions.*

Proof. Let E/k be solvable. Let F be a field containing k and assume E, F are subfields of some algebraically closed field. Let K be Galois solvable over k, and $E \subset K$. Then KF is Galois over F and $G(KF/F)$ is a subgroup of $G(K/k)$ by Theorem 1.12. Hence EF/F is solvable. It is clear that a subextension of a solvable extension is solvable. Let $E \supset F \supset k$ be a tower, and assume that E/F is solvable and F/k is solvable. Let K be a finite solvable Galois extension of k containing F. We just saw that EK/K is solvable. Let L be a solvable Galois extension of K containing EK. If σ is any embedding of L over k in a given algebraic closure, then $\sigma K = K$ and hence σL is a solvable extension of K. We let M be the compositum of all extensions σL for all embeddings σ of L over k.

Then M is Galois over k, and is therefore Galois over K. The Galois group of M over K is a subgroup of the product

$$\prod_{\sigma} G(\sigma L/K)$$

by Theorem 1.14. Hence it is solvable. We have a surjective homomorphism $G(M/k) \to G(K/k)$ by Theorem 1.10. Hence the Galois group of M/k has a solvable normal subgroup whose factor group is solvable. It is therefore solvable. Since $E \subset M$, our proof is complete.



A finite extension F of k is said to be **solvable by radicals** if it is separable and if there exists a finite extension E of k containing F, and admitting a tower decomposition

$$k = E_0 \subset E_1 \subset E_2 \subset \cdots \subset E_m = E$$

such that each step E_{i+1}/E_i is one of the following types:

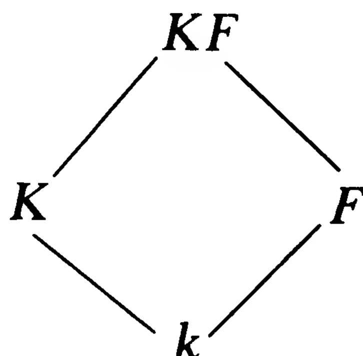
1. It is obtained by adjoining a root of unity.
2. It is obtained by adjoining a root of a polynomial $X^n - a$ with $a \in E_i$ and n prime to the characteristic.
3. It is obtained by adjoining a root of an equation $X^p - X - a$ with $a \in E_i$ if p is the characteristic > 0.

One can see at once that the class of extensions which are solvable by radicals is a distinguished class.

Theorem 7.2. *Let E be a separable extension of k. Then E is solvable by radicals if and only if E/k is solvable.*

Proof. Assume that E/k is solvable, and let K be a finite solvable Galois extension of k containing E. Let m be the product of all primes unequal to the characteristic dividing the degree $[K : k]$, and let $F = k(\zeta)$ where ζ is a primitive m-th root of unity. Then F/k is abelian. We lift K over F. Then KF is solvable over F. There is a tower of subfields between F and KF such that each step is cyclic of prime order, because every solvable group admits a tower of sub-

groups of the same type, and we can use Theorem 1.10. By Theorems 6.2 and 6.4, we conclude that KF is solvable by radicals over F, and hence is solvable by radicals over k. This proves that E/k is solvable by radicals.



Conversely, assume that E/k is solvable by radicals. For any embedding σ of E in E^a over k, the extension $\sigma E/k$ is also solvable by radicals. Hence the smallest Galois extension K of E containing k, which is a composite of E and its conjugates is solvable by radicals. Let m be the product of all primes unequal to the characteristic dividing the degree $[K : k]$ and again let $F = k(\zeta)$ where ζ is a primitive m-th root of unity. It will suffice to prove that KF is solvable over F, because it follows then that KF is solvable over k and hence $G(K/k)$ is solvable because it is a homomorphic image of $G(KF/k)$. But KF/F can be decomposed into a tower of extensions, such that each step is prime degree and of the type described in Theorem 6.2 or Theorem 6.4, and the corresponding root of unity is in the field F. Hence KF/F is solvable, and our theorem is proved.

Remark. One could modify our preceding discussion by not assuming separability. Then one must deal with normal extensions instead of Galois extensions, and one must allow equations $X^p - a$ in the solvability by radicals, with p equal to the characteristic. Then we still have the theorem corresponding to Theorem 7.2. The proof is clear in view of Chapter V, §6.

For a proof that every solvable group is a Galois group over the rationals, I refer to Shafarevich [Sh 54], as well as contributions of Iwasawa [Iw 53].

[Iw 53] K. Iwasawa, On solvable extension of algebraic number fields, *Ann. of Math.* **58** (1953), pp. 548–572

[Sh 54] I. Shafarevich, Construction of fields of algebraic numbers with given solvable Galois group, *Izv. Akad. Nauk SSSR* **18** (1954), pp. 525–578 (*Amer. Math. Soc. Transl.* **4** (1956), pp. 185–237)

§8. ABELIAN KUMMER THEORY

In this section we shall carry out a generalization of the theorem concerning cyclic extensions when the ground field contains enough roots of unity.

Let k be a field and m a positive integer. A Galois extension K of k with group G is said to be of **exponent** m if $\sigma^m = 1$ for all $\sigma \in G$.

We shall investigate abelian extensions of exponent m. We first assume that m is prime to the characteristic of k, and that k contains a primitive m-th root of unity. We denote by μ_m the group of m-th roots of unity. We assume that all our algebraic extensions in this section are contained in a fixed algebraic closure k^a.

Let $a \in k$. The symbol $a^{1/m}$ (or $\sqrt[m]{a}$) is not well defined. If $\alpha^m = a$ and ζ is an m-th root of unity, then $(\zeta\alpha)^m = a$ also. We shall use the symbol $a^{1/m}$ to denote any such element α, which will be called an m-th root of a. Since the roots of unity are in the ground field, we observe that the field $k(\alpha)$ is the same no matter which m-th root α of a we select. We denote this field by $k(a^{1/m})$.

We denote by k^{*m} the subgroup of k^* consisting of all m-th powers of non-zero elements of k. It is the image of k^* under the homomorphism $x \mapsto x^m$.

Let B be a subgroup of k^* containing k^{*m}. We denote by $k(B^{1/m})$ or K_B the composite of all fields $k(a^{1/m})$ with $a \in B$. It is uniquely determined by B as a subfield of k^a.

Let $a \in B$ and let α be an m-th root of a. The polynomial $X^m - a$ splits into linear factors in K_B, and thus K_B is Galois over k, because this holds for all $a \in B$. Let G be the Galois group. Let $\sigma \in G$. Then $\sigma\alpha = \omega_\sigma \alpha$ for some m-th root of unity $\omega_\sigma \in \mu_m \subset k^*$. The map

$$\sigma \mapsto \omega_\sigma$$

is obviously a homomorphism of G into μ_m, i.e. for $\tau, \sigma \in G$ we have

$$\tau\sigma\alpha = \omega_\tau \omega_\sigma \alpha = \omega_\sigma \omega_\tau \alpha.$$

We may write $\omega_\sigma = \sigma\alpha/\alpha$. This root of unity ω_σ is independent of the choice of m-th root of a, for if α' is another m-th root, then $\alpha' = \zeta\alpha$ for some $\zeta \in \mu_m$, whence

$$\sigma\alpha'/\alpha' = \zeta\sigma\alpha/\zeta\alpha = \sigma\alpha/\alpha.$$

We denote ω_σ by $\langle \sigma, a \rangle$. The map

$$(\sigma, a) \mapsto \langle \sigma, a \rangle$$

gives us a map

$$G \times B \to \mu_m.$$

If $a, b \in B$ and $\alpha^m = a$, $\beta^m = b$ then $(\alpha\beta)^m = ab$ and

$$\sigma(\alpha\beta)/\alpha\beta = (\sigma\alpha/\alpha)(\sigma\beta/\beta).$$

We conclude that the map above is bilinear. Furthermore, if $a \in k^{*m}$ it follows that $\langle \sigma, a \rangle = 1$.

Theorem 8.1. *Let k be a field, m an integer > 0 prime to the characteristic of k, and assume that a primitive m-th root of unity lies in k. Let B be a subgroup of k^* containing k^{*m} and let $K_B = k(B^{1/m})$. Then K_B is Galois, and abelian of exponent m. Let G be its Galois group. We have a bilinear map*

$$G \times B \to \mu_m \quad \textit{given by} \quad (\sigma, a) \mapsto \langle \sigma, a \rangle.$$

If $\sigma \in G$ and $a \in B$, and $\alpha^m = a$ then $\langle \sigma, a \rangle = \sigma\alpha/\alpha$. The kernel on the left is 1 and the kernel on the right is k^{*m}. The extension K_B/k is finite if and only if $(B : k^{*m})$ is finite. If that is the case, then

$$B/k^{*m} \approx G^{\wedge},$$

and in particular we have the equality

$$[K_B : k] = (B : k^{*m}).$$

Proof. Let $\sigma \in G$. Suppose $\langle \sigma, a \rangle = 1$ for all $a \in B$. Then for every generator α of K_B such that $\alpha^m = a \in B$ we have $\sigma\alpha = \alpha$. Hence σ induces the identity on K_B and the kernel on the left is 1. Let $a \in B$ and suppose $\langle \sigma, a \rangle = 1$ for all $\sigma \in G$. Consider the subfield $k(a^{1/m})$ of K_B. If $a^{1/m}$ is not in k, there exists an automorphism of $k(a^{1/m})$ over k which is not the identity. Extend this automorphism to K_B, and call this extension σ. Then clearly $\langle \sigma, a \rangle \neq 1$. This proves our contention.

By the duality theorem of Chapter I, §9 we see that G is finite if and only if B/k^{*m} is finite, and in that case we have the isomorphism as stated, so that in particular the order of G is equal to $(B : k^{*m})$, thereby proving the theorem.

Theorem 8.2. *Notation being as in Theorem 8.1, the map $B \mapsto K_B$ gives a bijection of the set of subgroups of k^* containing k^{*m} and the abelian extensions of k of exponent m.*

Proof. Let B_1, B_2 be subgroups of k^* containing k^{*m}. If $B_1 \subset B_2$ then $k(B_1^{1/m}) \subset k(B_2^{1/m})$. Conversely, assume that $k(B_1^{1/m}) \subset k(B_2^{1/m})$. We wish to prove $B_1 \subset B_2$. Let $b \in B_1$. Then $k(b^{1/m}) \subset k(B_2^{1/m})$ and $k(b^{1/m})$ is contained in a finitely generated subextension of $k(B_2^{1/m})$. Thus we may assume without loss of generality that B_2/k^{*m} is finitely generated, hence finite. Let B_3 be the subgroup of k^* generated by B_2 and b. Then $k(B_2^{1/m}) = k(B_3^{1/m})$ and from what we saw above, the degree of this field over k is precisely

$$(B_2 : k^{*m}) \quad \text{or} \quad (B_3 : k^{*m}).$$

Thus these two indices are equal, and $B_2 = B_3$. This proves that $B_1 \subset B_2$.

We now have obtained an injection of our set of groups B into the set of abelian extensions of k of exponent m. Assume finally that K is an abelian extension of k of exponent m. Any finite subextension is a composite of cyclic extensions of exponent m because any finite abelian group is a product of cyclic groups, and we can apply Corollary 1.16. By Theorem 6.2, every cyclic extension can be obtained by adjoining an m-th root. Hence K can be obtained by adjoining a family of m-th roots, say m-th roots of elements $\{b_j\}_{j \in J}$ with $b_j \in k^*$. Let B be the subgroup of k^* generated by all b_j and k^{*m}. If $b' = ba^m$ with $a, b \in k$ then obviously

$$k(b'^{1/m}) = k(b^{1/m}).$$

Hence $k(B^{1/m}) = K$, as desired.

When we deal with abelian extensions of exponent p equal to the characteristic, then we have to develop an additive theory, which bears the same relationship to Theorems 8.1 and 8.2 as Theorem 6.4 bears to Theorem 6.2.

If k is a field, we define the operator $\wp$ by

$$\wp(x) = x^p - x$$

for $x \in k$. Then $\wp$ is an additive homomorphism of k into itself. The subgroup $\wp(k)$ plays the same role as the subgroup k^{*m} in the multiplicative theory, whenever m is a prime number. The theory concerning a power of p is slightly more elaborate and is due to Witt.

We now assume k has characteristic p. A root of the polynomial $X^p - X - a$ with $a \in k$ will be denoted by $\wp^{-1}a$. If B is a subgroup of k containing $\wp k$ we let $K_B = k(\wp^{-1}B)$ be the field obtained by adjoining $\wp^{-1}a$ to k for all $a \in B$. We emphasize the fact that B is an *additive* subgroup of k.

Theorem 8.3. *Let k be a field of characteristic p. The map $B \mapsto k(\wp^{-1}B)$ is a bijection between subgroups of k containing $\wp k$ and abelian extensions of k of exponent p. Let $K = K_B = k(\wp^{-1}B)$, and let G be its Galois group. If $\sigma \in G$ and $a \in B$, and $\wp\alpha = a$, let $\langle \sigma, a \rangle = \sigma\alpha - \alpha$. Then we have a bilinear map*

$$G \times B \to \mathbf{Z}/p\mathbf{Z} \quad \textit{given by} \quad (\sigma, a) \to \langle \sigma, a \rangle.$$

The kernel on the left is 1 and the kernel on the right is $\wp k$. The extension K_B/k is finite if and only if $(B : \wp k)$ is finite and if that is the case, then

$$[K_B : k] = (B : \wp k).$$

Proof. The proof is entirely similar to the proof of Theorems 8.1 and 8.2. It can be obtained by replacing multiplication by addition, and using the "$\wp$-th root" instead of an m-th root. Otherwise, there is no change in the wording of the proof.

The analogous theorem for abelian extensions of exponent p^n requires Witt vectors, and will be developed in the exercises.

Bibliography

[Wi 35] E. WITT, Der Existenzsatz für abelsche Funktionenkörper, *J. reine angew. Math.* **173** (1935), pp. 43–51

[Wi 36] E. WITT, Konstruktion von galoisschen Körpern der Charakteristik p mit vorgegebener Gruppe der Ordung p^f, *J. reine angew. Math.* **174** (1936), pp. 237–245

[Wi 37] E. WITT, Zyklische Körper und Algebren der Charakteristik p vom Grad p^n. Struktur diskret bewerteter perfekter Körper mit vollkommenem Restklassenkörper der Charakteristik p, *J. reine angew. Math.* **176** (1937), pp. 126–140

§9. THE EQUATION $X^n - a = 0$

When the roots of unity are not in the ground field, the equation $X^n - a = 0$ is still interesting but a little more subtle to treat.

Theorem 9.1. *Let k be a field and n an integer $\geqq 2$. Let $a \in k, a \neq 0$. Assume that for all prime numbers p such that $p|n$ we have $a \notin k^p$, and if $4|n$ then $a \notin -4k^4$. Then $X^n - a$ is irreducible in $k[X]$.*

Proof. Our first assumption means that a is not a p-th power in k. We shall reduce our theorem to the case when n is a prime power, by induction.

Write $n = p^r m$ with p prime to m, and p odd. Let

$$X^m - a = \prod_{\nu=1}^{m} (X - \alpha_\nu)$$

be the factorization of $X^m - a$ into linear factors, and say $\alpha = \alpha_1$. Substituting X^{p^r} for X we get

$$X^n - a = X^{p^r m} - a = \prod_{\nu=1}^{m} (X^{p^r} - \alpha_\nu).$$

We may assume inductively that $X^m - a$ is irreducible in $k[X]$. We contend that α is not a p-th power in $k(\alpha)$. Otherwise, $\alpha = \beta^p$, $\beta \in k(\alpha)$. Let N be the norm from $k(\alpha)$ to k. Then

$$-a = (-1)^m N(\alpha) = (-1)^m N(\beta^p) = (-1)^m N(\beta)^p.$$

If m is odd, a is a p-th power, which is impossible. Similarly, if m is even and p is odd, we also get a contradiction. This proves our contention, because m is prime to p. If we know our theorem for prime powers, then we conclude that $X^{p^r} - \alpha$ is irreducible over $k(\alpha)$. If A is a root of $X^{p^r} - \alpha$ then $k \subset k(\alpha) \subset k(A)$ gives a tower, of which the bottom step has degree m and the top step has degree p^r. It follows that A has degree n over k and hence that $X^n - a$ is irreducible.

We now suppose that $n = p^r$ is a prime power.

If p is the characteristic, let α be a p-th root of a. Then $X^p - a = (X - \alpha)^p$ and hence $X^{p^r} - a = (X^{p^{r-1}} - \alpha)^p$ if $r \geqq 2$. By an argument even more trivial than before, we see that α is not a p-th power in $k(\alpha)$, hence inductively $X^{p^{r-1}} - \alpha$ is irreducible over $k(\alpha)$. Hence $X^{p^r} - a$ is irreducible over k.

Suppose that p is not the characteristic. We work inductively again, and let α be a root of $X^p - a$.

Suppose a is not a p-th power in k. We claim that $X^p - a$ is irreducible. Otherwise a root α of $X^p - a$ generates an extension $k(\alpha)$ of degree $d < p$ and $\alpha^p = a$. Taking the norm from $k(\alpha)$ to k we get $N(\alpha)^p = a^d$. Since d is prime to p, it follows that α is a p-th power in k, contradiction.

Let $r \geqq 2$. We let $\alpha = \alpha_1$. We have

$$X^p - a = \prod_{\nu=1}^{p} (X - \alpha_\nu)$$

and

$$X^{p^r} - a = \prod_{\nu=1}^{p} (X^{p^{r-1}} - \alpha_\nu).$$

Assume that α is not a p-th power in $k(\alpha)$. Let A be a root of $X^{p^{r-1}} - \alpha$. If p is odd then by induction, A has degree p^{r-1} over $k(\alpha)$, hence has degree p^r over k and we are done. If $p = 2$, suppose $\alpha = -4\beta^4$ with $\beta \in k(\alpha)$. Let N be the norm from $k(\alpha)$ to k. Then $-a = N(\alpha) = 16N(\beta)^4$, so $-a$ is a square in k. Since $p = 2$ we get $\sqrt{-1} \in k(\alpha)$ and $\alpha = (\sqrt{-1}\, 2\beta^2)^2$, a contradiction. Hence again by induction, we find that A has degree p^r over k. We therefore assume that $\alpha = \beta^p$ with some $\beta \in k(\alpha)$, and derive the consequences.

Taking the norm from $k(\alpha)$ to k we find

$$-a = (-1)^p N(\alpha) = (-1)^p N(\beta^p) = (-1)^p N(\beta)^p.$$

If p is odd, then a is a p-th power in k, contradiction. Hence $p = 2$, and

$$-a = N(\beta)^2$$

is a square in k. Write $-a = b^2$ with $b \in k$. Since a is not a square in k we conclude that -1 is not a square in k. Let $i^2 = -1$. Over $k(i)$ we have the factorization

$$X^{2^r} - a = X^{2^r} + b^2 = (X^{2^{r-1}} + ib)(X^{2^{r-1}} - ib).$$

Each factor is of degree 2^{r-1} and we argue inductively. If $X^{2^{r-1}} \pm ib$ is reducible over $k(i)$ then $\pm ib$ is a square in $k(i)$ or lies in $-4(k(i))^4$. In either case, $\pm ib$ is a square in $k(i)$, say

$$\pm ib = (c + di)^2 = c^2 + 2cdi - d^2$$

with $c, d \in k$. We conclude that $c^2 = d^2$ or $c = \pm d$, and $\pm ib = 2cdi = \pm 2c^2 i$. Squaring gives a contradiction, namely

$$a = -b^2 = -4c^4.$$

We now conclude by unique factorization that $X^{2^r} + b^2$ cannot factor in $k[X]$, thereby proving our theorem.

The conditions of our theorem are necessary because

$$X^4 + 4b^4 = (X^2 + 2bX + 2b^2)(X^2 - 2bX + 2b^2).$$

If $n = 4m$ and $a \in -4k^4$ then $X^n - a$ is reducible.

Corollary 9.2. *Let k be a field and assume that $a \in k$, $a \neq 0$, and that a is not a p-th power for some prime p. If p is equal to the characteristic, or if p is odd, then for every integer $r \geqq 1$ the polynomial $X^{p^r} - a$ is irreducible over k.*

Proof. The assertion is logically weaker than the assertion of the theorem.

Corollary 9.3. *Let k be a field and assume that the algebraic closure k^a of k is of finite degree > 1 over k. Then $k^a = k(i)$ where $i^2 = -1$, and k has characteristic 0.*

Proof. We note that k^a is normal over k. If k^a is not separable over k, so char $k = p > 0$, then k^a is purely inseparable over some subfield of degree $>$ 1 (by Chapter V, §6), and hence there is a subfield E containing k, and an element $a \in E$ such that $X^p - a$ is irreducible over E. By Corollary 9.2, k^a cannot be of finite degree over E. (The reader may restrict his or her attention to characteristic 0 if Chapter V, §6 was omitted.)

We may therefore assume that k^a is Galois over k. Let $k_1 = k(i)$. Then k^a is also Galois over k_1. Let G be the Galois group of k^a/k_1. Suppose that there is a prime number p dividing the order of G, and let H be a subgroup of order p. Let F be its fixed field. Then $[k^a : F] = p$. If p is the characteristic, then Exercise 29 at the end of the chapter will give the contradiction. We may assume that p is not the characteristic. The p-th roots of unity $\neq 1$ are the roots of a polynomial of degree $\leqq p - 1$ (namely $X^{p-1} + \cdots + 1$), and hence must lie in F. By Theorem 6.2, it follows that k^a is the splitting field of some polynomial $X^p - a$ with $a \in F$. The polynomial $X^{p^2} - a$ is necessarily reducible. By the theorem, we must have $p = 2$ and $a = -4b^4$ with $b \in F$. This implies

$$k^a = F(a^{1/2}) = F(i).$$

But we assumed $i \in k_1$, contradiction.

Thus we have proved $k^a = k(i)$. It remains to prove that char $k = 0$, and for this I use an argument shown to me by Keith Conrad. We first show that a sum of squares in k is a square. It suffices to prove this for a sum of two squares, and in this case we write an element $x + iy \in k(i) = k^a$ as a square.

$$x + iy = (u + iv)^2, \qquad x, y, u, v \in k,$$

and then $x^2 + y^2 = (u^2 + v^2)^2$. Then to prove k has characteristic 0, we merely observe that if the characteristic is > 0, then -1 is a finite sum $1 + \ldots + 1$, whence a square by what we have just shown, but $k^a = k(i)$, so this concludes the proof.

Corollary 9.3 is due to Artin; see [Ar 24], given at the end of Chapter XI. In that chapter, much more will be proved about the field k.

Example 1. Let $k = \mathbf{Q}$ and let $G_{\mathbf{Q}} = G(\mathbf{Q}^a/\mathbf{Q})$. Then the only non-trivial torsion elements in $G_{\mathbf{Q}}$ have order 2. It follows from Artin's theory (as given in Chapter XI) that all such torsion elements are conjugate in $G_{\mathbf{Q}}$. One uses Chapter XI, Theorems 2.2, 2.4, and 2.9.)

Example 2. Let k be a field of characteristic not dividing n. Let $a \in k$, $a \neq 0$ and let K be the splitting field of $X^n - a$. Let α be one root of $X^n - a$, and let ζ be a primitive n-th root of unity. Then

$$K = k(\alpha, \zeta) = k(\alpha, \boldsymbol{\mu}_n).$$

We assume the reader is acquainted with matrices over a commutative ring. Let $\sigma \in G_{K/k}$. Then $(\sigma\alpha)^n = a$, so there exists some integer $b = b(\sigma)$ uniquely determined mod n, such that

$$\sigma(\alpha) = \alpha\zeta^{b(\sigma)}.$$

Since σ induces an automorphism of the cyclic group $\boldsymbol{\mu}_n$, there exists an integer $d(\sigma)$ relatively prime to n and uniquely determined mod n such that $\sigma(\zeta) = \zeta^{d(\sigma)}$. Let $G(n)$ be the subgroup of $GL_2(\mathbf{Z}/n\mathbf{Z})$ consisting of all matrices

$$M = \begin{pmatrix} 1 & 0 \\ b & d \end{pmatrix} \text{ with } b \in \mathbf{Z}/n\mathbf{Z} \quad \text{and} \quad d \in (\mathbf{Z}/n\mathbf{Z})^*.$$

Observe that $\#G(n) = n\varphi(n)$. We obtain an injective map

$$\sigma \mapsto M(\sigma) = \begin{pmatrix} 1 & 0 \\ b(\sigma) & d(\sigma) \end{pmatrix} \quad \text{of } G_{K/k} \hookrightarrow G(n),$$

which is immediately verified to be an injective homomorphism. The question arises, when is it an isomorphism? The next theorem gives an answer over some fields, applicable especially to the rational numbers.

Theorem 9.4. *Let k be a field. Let n be an odd positive integer prime to the characteristic, and assume that $[k(\boldsymbol{\mu}_n) : k] = \varphi(n)$. Let $a \in k$, and suppose that for each prime $p|n$ the element a is not a p-th power in k. Let K be the splitting field of $X^n - a$ over k. Then the above homomorphism $\sigma \mapsto M(\sigma)$ is an isomorphism of $G_{K/k}$ with $G(n)$. The commutator group is* $\mathrm{Gal}(K/k(\boldsymbol{\mu}_n))$, *so $k(\boldsymbol{\mu}_n)$ is the maximal abelian subextension of K.*

Proof. This is a special case of the general theory of §11, and Exercise 39, taking into account the representation of $G_{K/k}$ in the group of matrices. One need only use the fact that the order of $G_{K/k}$ is $n\varphi(n)$, according to that exercise, and so $\#(G_{K/k}) = \#G(n)$, so $G_{K/k} = G(n)$. However, we shall given an independent proof as an example of techniques of Galois theory. We prove the theorem by induction.

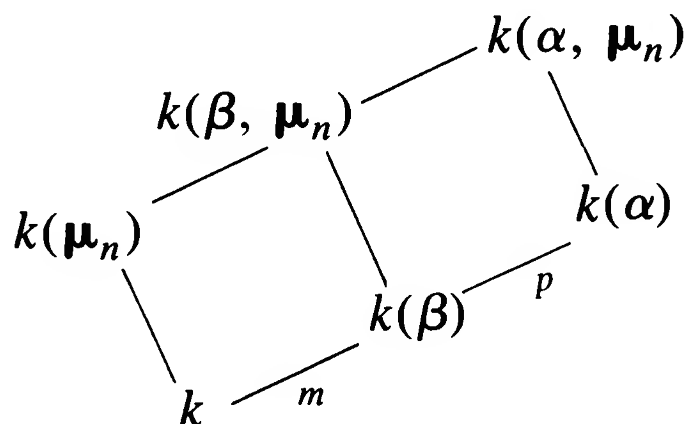
Suppose first $n = p$ is prime. Since $[k(\boldsymbol{\mu}_p) : k] = p - 1$ is prime to p, it follows that if α is a root of $X^p - a$, then $k(\alpha) \cap k(\boldsymbol{\mu}_p) = k$ because $[k(\alpha) : k] = p$. Hence $[K : k] = p(p - 1)$, so $G_{K/k} = G(p)$.

A direct computation of a commutator of elements in $G(n)$ for arbitrary n shows that the commutator subgroup is contained in the group of matrices

$$\begin{pmatrix} 1 & 0 \\ b & 1 \end{pmatrix}, \ b \in \mathbf{Z}/n\mathbf{Z},$$

and so must be that subgroup because its factor group is isomorphic to $(\mathbf{Z}/n\mathbf{Z})^*$ under the projection on the diagonal. This proves the theorem when $n = p$.

Now let $p \mid n$ and write $n = pm$. Then $[k(\boldsymbol{\mu}_m) : k] = \varphi(m)$, immediately from the hypothesis that $[k(\boldsymbol{\mu}_n) : k] = \varphi(n)$. Let α be a root of $X^n - a$, and let $\beta = \alpha^p$. Then β is a root of $X^m - a$, and by induction we can apply the theorem to $X^m - a$. The field diagram is as follows.



Since α has degree pm over k, it follows that α cannot have lower degree than p over $k(\beta)$, so $[k(\alpha) : k(\beta)] = p$ and $X^p - \beta$ is irreducible over $k(\beta)$. We apply the first part of the proof to $X^p - \beta$ over $k(\beta)$. The property concerning the maximal abelian subextension of the splitting field shows that

$$k(\alpha) \cap k(\beta, \boldsymbol{\mu}_n) = k(\beta).$$

Hence $[k(\alpha, \boldsymbol{\mu}_n) : k(\beta, \boldsymbol{\mu}_n)] = p$. By induction, $[k(\beta, \boldsymbol{\mu}_n) : k(\boldsymbol{\mu}_n)] = m$, again because of the maximal abelian subextension of the splitting field of $X^m - a$ over k. This proves that $[K : k] = n\varphi(n)$, whence $G_{K/k} = G(n)$, and the commutator statement has already been proved. This concludes the proof of Theorem 9.4.

Remarks. When n is even, there are some complications, because for instance $\mathbf{Q}(\sqrt{2})$ is contained in $\mathbf{Q}(\boldsymbol{\mu}_8)$, so there are dependence relations among the fields in question. The non-abelian extensions, as in Theorem 9.4, are of intrinsic interest because they constitute the first examples of such extensions that come to mind, but they arose in other important contexts. For instance, Artin used them to give a probabilistic model for the density of primes p such that 2 (say) is a primitive root mod p (that is, 2 generates the cyclic group $(\mathbf{Z}/p\mathbf{Z})^*$. Instead of 2 he took any non-square integer $\neq \pm 1$. At first, Artin did not realize explicitly the above type of dependence, and so came to an answer that was off by some factor in some cases. Lehmer discovered the discrepancy by computations. As Artin then said, one has to multiply by the "obvious" factor which reflects the field dependencies. Artin never published his conjecture, but the matter is discussed in detail by Lang-Tate in the introduction to his collected papers (Addison-Wesley, Springer Verlag).

Similar conjectural probabilistic models were constructed by Lang-Trotter in connection with elliptic curves, and more generally with certain p-adic representations of the Galois group, in "Primitive points on elliptic curves", *Bull. AMS* **83** No. 2 (1977), pp. 289–292; and [LaT 75] (end of §14).

For further comments on the p-adic representations of Galois groups, see §14 and §15.

§10. GALOIS COHOMOLOGY

Let G be a group and A an abelian group which we write additively for the general remarks which we make, preceding our theorems. Let us assume that G operates on A, by means of a homomorphism $G \to \mathrm{Aut}(A)$. By a **1-cocycle** of G in A one means a family of elements $\{\alpha_\sigma\}_{\sigma \in G}$ with $\alpha_\sigma \in A$, satisfying the relations

$$\alpha_\sigma + \sigma\alpha_\tau = \alpha_{\sigma\tau}$$

for all $\sigma, \tau \in G$. If $\{\alpha_\sigma\}_{\sigma \in G}$ and $\{\beta_\sigma\}_{\sigma \in G}$ are 1-cocycles, then we can add them to get a 1-cocycle $\{\alpha_\sigma + \beta_\sigma\}_{\sigma \in G}$. It is then clear that 1-cocycles form a group, denoted by $Z^1(G, A)$. By a **1-coboundary** of G in A one means a family of elements $\{\alpha_\sigma\}_{\sigma \in G}$ such that there exists an element $\beta \in A$ for which $\alpha_\sigma = \sigma\beta - \beta$ for all $\sigma \in G$. It is then clear that a 1-coboundary is a 1-cocycle, and that the 1-coboundaries form a group, denoted by $B^1(G, A)$. The factor group

$$Z^1(G, A)/B^1(G, A)$$

is called the **first cohomology group** of G in A and is denoted by $H^1(G, A)$.

Remarks. Suppose G is cyclic. Let

$$\mathrm{Tr}_G\colon A \to A \text{ be the homomorphism } a \mapsto \sum_{\sigma \in G} \sigma(a).$$

Let γ be a generator of G. Let $(1 - \gamma)A$ be the subgroup of A consisting of all elements $a - \gamma(a)$ with $a \in A$. Then $(1 - \gamma)A$ is contained in $\ker \mathrm{Tr}_G$. The reader will verify as an exercise that there is an isomorphism

$$\ker \mathrm{Tr}_G/(1 - \gamma)A \approx H^1(G, A).$$

Then the next theorem for a cyclic group is just Hilbert's Theorem 90 of §6. *Cf.* also the cohomology of groups, Chapter XX, Exercise 4, for an even more general context.

Theorem 10.1. *Let K/k be a finite Galois extension with Galois group G. Then for the operation of G on K^* we have $H^1(G, K^*) = 1$, and for the operation of G on the additive group of K we have $H^1(G, K) = 0$. In other words, the first cohomology group is trivial in both cases.*

Proof. Let $\{\alpha_\sigma\}_{\sigma \in G}$ be a 1-cocycle of G in K^*. The multiplicative cocycle relation reads

$$\alpha_\sigma \alpha_\tau^\sigma = \alpha_{\sigma\tau}.$$

By the linear independence of characters, there exists $\theta \in K$ such that the element

$$\beta = \sum_{\tau \in G} \alpha_\tau \tau(\theta)$$

is $\neq 0$. Then

$$\sigma\beta = \sum_{\tau \in G} \alpha_\tau^\sigma \sigma\tau(\theta) = \sum_{\tau \in G} \alpha_{\sigma\tau}\alpha_\sigma^{-1}\sigma\tau(\theta)$$

$$= \alpha_\sigma^{-1} \sum_{\tau \in G} \alpha_{\sigma\tau}\sigma\tau(\theta) = \alpha_\sigma^{-1}\beta.$$

We get $\alpha_\sigma = \beta/\sigma\beta$, and using β^{-1} instead of β gives what we want.

For the additive part of the theorem, we find an element $\theta \in K$ such that the trace $\mathrm{Tr}(\theta)$ is not equal to 0. Given a 1-cocycle $\{\alpha_\sigma\}$ in the additive group of K, we let

$$\beta = \frac{1}{\mathrm{Tr}(\theta)} \sum_{\tau \in G} \alpha_\tau \tau(\theta).$$

It follows at once that $\alpha_\sigma = \beta - \sigma\beta$, as desired.

The next lemma will be applied to the non-abelian Kummer theory of the next section.

Lemma 10.2. (Sah). *Let G be a group and let E be a G-module. Let τ be in the center of G. Then $H^1(G, E)$ is annihilated by the map $x \mapsto \tau x - x$ on E. In particular, if this map is an automorphism of E, then $H^1(G, E) = 0$.*

Proof. Let f be a 1-cocycle of G in E. Then

$$f(\sigma) = f(\tau\sigma\tau^{-1}) = f(\tau) + \tau(f(\sigma\tau^{-1})$$
$$= f(\tau) + \tau[f(\sigma) + \sigma f(\tau^{-1})].$$

Therefore

$$\tau f(\sigma) - f(\sigma) = -\sigma\tau f(\tau^{-1}) - f(\tau).$$

But $f(1) = f(1) + f(1)$ implies $f(1) = 0$, and

$$0 = f(1) = f(\tau\tau^{-1}) = f(\tau) + \tau f(\tau^{-1}).$$

This shows that $(\tau - 1)f(\sigma) = (\sigma - 1)f(\tau)$, so f is a coboundary. This proves the lemma.

§11. NON-ABELIAN KUMMER EXTENSIONS

We are interested in the splitting fields of equations $X^n - a = 0$ when the n-th roots of unity are not contained in the ground field. More generally, we want to know roughly (or as precisely as possible) the Galois group of simultaneous equations of this type. For this purpose, we axiomatize the pattern of proof to an additive notation, which in fact makes it easier to see what is going on.

We fix an integer $N > 1$, and we let M range over positive integers dividing N. We let P be the set of primes dividing N. We let G be a group, and let:

A = G-module such that the isotropy group of any element of A is of finite index in G. We also assume that A is divisible by the primes $p|N$, that is

$$pA = A \qquad \text{for all } p \in P.$$

Γ = finitely generated subgroup of A such that Γ is pointwise fixed by G.

We assume that A_N is finite. Then $\frac{1}{N}\Gamma$ is also finitely generated. Note that

$$\frac{1}{N}\Gamma \supset A_N.$$

Example. For our purposes here, the above situation summarizes the properties which hold in the following situation. Let K be a finitely generated field over the rational numbers, or even a finite extension of the rational numbers. We let A be the multiplicative group of the algebraic closure K^a. We let $G = G_K$ be the Galois group $\text{Gal}(K^a/K)$. We let Γ be a finitely generated subgroup of the multiplicative group K^*. Then all the above properties are satisfied. We see that $A_N = \mu_N$ is the group of N-th roots of unity. The group written $\frac{1}{N}\Gamma$ in additive notation is written $\Gamma^{1/N}$ in multiplicative notation.

Next we define the appropriate groups analogous to the Galois groups of Kummer theory, as follows. For any G-submodule B of A, we let:

$G(B)$ = image of G in $\text{Aut}(B)$,

$G(N) = G(A_N)$ = image of G in $\text{Aut}(A_N)$,

$H(N)$ = subgroup of G leaving A_N pointwise fixed,

$H_\Gamma(M, N)$ (for $M|N$) = image of $H(N)$ in $\text{Aut}\left(\frac{1}{M}\Gamma\right)$.

Then we have an exact sequence:

$$0 \to H_\Gamma(M, N) \to G\left(\frac{1}{M}\Gamma + A_N\right) \to G(N) \to 0.$$

Example. In the concrete case mentioned above, the reader will easily recognize these various groups as Galois groups. For instance, let A be the multiplicative group. Then we have the following lattice of field extensions with corresponding Galois groups:

$$G(\Gamma^{1/M}\mu_N)\left\{\begin{array}{c} K(\mu_N, \Gamma^{1/M}) \\ | \\ K(\mu_N) \\ | \\ K \end{array}\right. \begin{array}{l} \left.\vphantom{\begin{array}{c}a\\b\end{array}}\right\} H_\Gamma(M, N) \\ \left.\vphantom{\begin{array}{c}a\\b\end{array}}\right\} G(N) \end{array}$$

In applications, we want to know how much degeneracy there is when we translate $K(\mu_M, \Gamma^{1/M})$ over $K(\mu_N)$ with $M \mid N$. This is the reason we play with the pair M, N rather than a single N.

Let us return to a general Kummer representation as above. We are interested especially in that part of $(\mathbf{Z}/N\mathbf{Z})^*$ contained in $G(N)$, namely the group of integers n (mod N) such that there is an element $[n]$ in $G(N)$ such that

$$[n]a = na \qquad \text{for all } a \in A_N.$$

Such elements are always contained in the center of $G(N)$, and are called **homotheties**.

Write

$$N = \prod p^{n(p)}$$

Let S be a subset of P. We want to make some non-degeneracy assumptions about $G(N)$. We call S the **special set**.

There is a product decomposition

$$(\mathbf{Z}/N\mathbf{Z})^* = \prod_{p \mid N} (\mathbf{Z}/p^{n(p)}\mathbf{Z})^*.$$

If $2 \mid N$ we suppose that $2 \in S$. For each $p \in S$ we suppose that there is an integer $c(p) = p^{f(p)}$ with $f(p) \geqq 1$ such that

$$G(A_N) \supset \prod_{p \in S} U_{c(p)} \times \prod_{p \notin S} (\mathbf{Z}/p^{n(p)}\mathbf{Z})^*,$$

where $U_{c(p)}$ is the subgroup of $\mathbf{Z}(p^{n(p)})$ consisting of those elements $\equiv 1 \bmod c(p)$.

The product decomposition on the right is relative to the direct sum decomposition

$$A_N = \bigoplus_{p|N} A_{p^{n(p)}}.$$

The above assumption will be called the non-degeneracy assumption. The integers $c(p)$ measure the extent to which $G(A_N)$ is degenerate.

Under this assumption, we observe that

$$[2] \in G(A_M) \quad \text{if} \quad M|N \text{ and } M \text{ is not divisible by primes of } S;$$

$$[1 + c] \in G(A_M) \quad \text{if} \quad M|N \text{ and } M \text{ is divisible only by primes of } S,$$

where

$$c = c(S) = \prod_{p \in S} c(p).$$

We can then use $[2] - [1] = [1]$ and $[1 + c] - [1] = [c]$ in the context of Lemma 10.2, since $[1]$ and $[c]$ are in the center of G.

For any M we define

$$c(M) = \prod_{\substack{p|M \\ p \in S}} c(p).$$

Define

$$\Gamma' = \frac{1}{N}\Gamma \cap A^G$$

and the **exponent**

$$e(\Gamma'/\Gamma) = \text{smallest positive integer } e \text{ such that } e\Gamma' \subset \Gamma.$$

It is clear that degeneracy in the Galois group $H_\Gamma(M, N)$ defined above can arise from lots of roots of unity in the ground field, or at least degeneracy in the Galois group of roots of unity; and also if we look at an equation

$$X^M - a = 0,$$

from the fact that a is already highly divisible in K. This second degeneracy would arise from the exponent $e(\Gamma'/\Gamma)$, as can be seen by looking at the Galois group of the divisions of Γ. The next theorem shows that these are the only sources of degeneracy.

We have the abelian Kummer pairing for $M|N$,

$$H_\Gamma(M, N) \times \Gamma/M\Gamma \to A_M \quad \text{given by} \quad (\tau, x) \mapsto \tau y - y,$$

where y is any element such that $My = x$. The value of the pairing is indepen-

dent of the choice of y. Thus for $x \in \Gamma$, we have a homomorphism

$$\varphi_x : H_\Gamma(M, N) \to A_M$$

such that

$$\varphi_x(\tau) = \tau y - y, \qquad \text{where } My = x.$$

Theorem 11.1. *Let $M \mid N$. Let φ be the homomorphism*

$$\varphi : \Gamma \to \operatorname{Hom}(H_\Gamma(M, N), A_M)$$

and let Γ_φ be its kernel. Let $e_M(\Gamma) =$ g.c.d. $(e(\Gamma'/\Gamma), M)$. Under the non-degeneracy assumption, we have

$$c(M)e_M(\Gamma)\Gamma_\varphi \subset M\Gamma.$$

Proof. Let $x \in \Gamma$ and suppose $\varphi_x = 0$. Let $My = x$. For $\sigma \in G$ let

$$y_\sigma = \sigma y - y.$$

Then $\{y_\sigma\}$ is a 1-cocycle of G in A_M, and by the hypothesis that $\varphi_x = 0$, this cocycle depends only on the class of σ modulo the subgroup of G leaving the elements of A_N fixed. In other words, we may view $\{y_\sigma\}$ as a cocycle of $G(N)$ in A_M. Let $c = c(N)$. By Lemma 10.2, it follows that $\{cy_\sigma\}$ splits as a cocycle of $G(N)$ in A_M. In other words, there exists $t_0 \in A_M$ such that

$$cy_\sigma = \sigma t_0 - t_0,$$

and this equation in fact holds for $\sigma \in G$. Let t be such that $ct = t_0$. Then

$$c\sigma y - cy = \sigma ct - cy,$$

whence $c(y - t)$ is fixed by all $\sigma \in G$, and therefore lies in $\frac{1}{N}\Gamma$. Therefore

$$e(\Gamma'/\Gamma)c(y - t) \in \Gamma.$$

We multiply both sides by M and observe that $cM(y - t) = cMy = cx$. This shows that

$$c(N)e(\Gamma'/\Gamma)\Gamma_\varphi \subset M\Gamma.$$

Since $\Gamma/M\Gamma$ has exponent M, we may replace $e(\Gamma'/\Gamma)$ by the greatest common divisor as stated in the theorem, and we can replace $c(N)$ by $c(M)$ to conclude the proof.

Corollary 11.2. *Assume that M is prime to $2(\Gamma' : \Gamma)$ and is not divisible by any primes of the special set S. Then we have an injection*

$$\varphi : \Gamma/M\Gamma \to \operatorname{Hom}(H_\Gamma(M, N), A_M).$$

If in addition Γ is free with basis $\{a_1, \ldots, a_r\}$, and we let $\varphi_i = \varphi_{a_i}$, then the map

$$H_\Gamma(M, N) \to A_M^{(r)} \quad \text{given by} \quad \tau \to (\varphi_1(\tau), \ldots, \varphi_r(\tau))$$

is injective. If A_M is cyclic of order M, this map is an isomorphism.

Proof. Under the hypotheses of the corollary, we have $c(M) = 1$ and $c_M(\Gamma) = 1$ in the theorem.

Example. Consider the case of Galois theory when A is the multiplicative group of K^a. Let $a_1, \ldots, a_r$ be elements of K^* which are multiplicatively independent. They generate a group as in the corollary. Furthermore, $A_M = \boldsymbol{\mu}_M$ is cyclic, so the corollary applies. If M is prime to $2(\Gamma' : \Gamma)$ and is not divisible by any primes of the special set S, *we have an isomorphism*

$$\varphi : \Gamma/M\Gamma \to \mathrm{Hom}(H_\Gamma(M, N), \boldsymbol{\mu}_M).$$

§12. ALGEBRAIC INDEPENDENCE OF HOMOMORPHISMS

Let A be an additive group, and let K be a field. Let $\lambda_1, \ldots, \lambda_n : A \to K$ be additive homomorphisms. We shall say that $\lambda_1, \ldots, \lambda_n$ are **algebraically dependent** (over K) if there exists a polynomial $f(X_1, \ldots, X_n)$ in $K[X_1, \ldots, X_n]$ such that for all $x \in A$ we have

$$f(\lambda_1(x), \ldots, \lambda_n(x)) = 0,$$

but such that f does not induce the zero function on $K^{(n)}$, i.e. on the direct product of K with itself n times. We know that with each polynomial we can associate a unique reduced polynomial giving the same function. If K is infinite, the reduced polynomial is equal to f itself. In our definition of dependence, we could as well assume that f is reduced.

A polynomial $f(X_1, \ldots, X_n)$ will be called **additive** if it induces an additive homomorphism of $K^{(n)}$ into K. Let $(Y) = (Y_1, \ldots, Y_n)$ be variables independent from (X). Let

$$g(X, Y) = f(X + Y) - f(X) - f(Y)$$

where $X + Y$ is the componentwise vector addition. Then the total degree of g viewed as a polynomial in (X) with coefficients in $K[Y]$ is strictly less than the total degree of f, and similarly, its degree in each X_i is strictly less than the degree of f in each X_i. One sees this easily by considering the difference of monomials,

$$M_{(\nu)}(X + Y) - M_{(\nu)}(X) - M_{(\nu)}(Y)$$
$$= (X_1 + Y_1)^{\nu_1} \cdots (X_n + Y_n)^{\nu_n} - X_1^{\nu_1} \cdots X_n^{\nu_n} - Y_1^{\nu_1} \cdots Y_n^{\nu_n}.$$

A similar assertion holds for g viewed as a polynomial in (Y) with coefficients in $K[X]$.

If f is reduced, it follows that g is reduced. Hence if f is additive, it follows that g is the zero polynomial.

Example. Let K have characteristic p. Then in one variable, the map

$$\xi \mapsto a\xi^{p^m}$$

for $a \in K$ and $m \geqq 1$ is additive, and given by the additive polynomial aX^{p^m}. We shall see later that this is a typical example.

Theorem 12.1. (Artin). *Let $\lambda_1, \ldots, \lambda_n : A \to K$ be additive homomorphisms of an additive group into a field. If these homomorphisms are algebraically dependent over K, then there exists an additive polynomial*

$$f(X_1, \ldots, X_n) \neq 0$$

in $K[X]$ such that

$$f(\lambda_1(x), \ldots, \lambda_n(x)) = 0$$

for all $x \in A$.

Proof. Let $f(X) = f(X_1, \ldots, X_n) \in K[X]$ be a reduced polynomial of lowest possible degree such that $f \neq 0$ but for all $x \in A$, $f(\Lambda(x)) = 0$, where $\Lambda(x)$ is the vector $(\lambda_1(x), \ldots, \lambda_n(x))$. We shall prove that f is additive.

Let $g(X, Y) = f(X + Y) - f(X) - f(Y)$. Then

$$g(\Lambda(x), \Lambda(y)) = f(\Lambda(x + y)) - f(\Lambda(x)) - f(\Lambda(y)) = 0$$

for all $x, y \in A$. We shall prove that g induces the zero function on $K^{(n)} \times K^{(n)}$. Assume otherwise. We have two cases.

Case 1. We have $g(\xi, \Lambda(y)) = 0$ for all $\xi \in K^{(n)}$ and all $y \in A$. By hypothesis, there exists $\xi' \in K^{(n)}$ such that $g(\xi', Y)$ is not identically 0. Let $P(Y) = g(\xi', Y)$. Since the degree of g in (Y) is strictly smaller than the degree of f, we have a contradiction.

Case 2. There exist $\xi' \in K^{(n)}$ and $y' \in A$ such that $g(\xi', \Lambda(y')) \neq 0$. Let $P(X) = g(X, \Lambda(y'))$. Then P is not the zero polynomial, but $P(\Lambda(x)) = 0$ for all $x \in A$, again a contradiction.

We conclude that g induces the zero function on $K^{(n)} \times K^{(n)}$, which proves what we wanted, namely that f is additive.

We now consider additive polynomials more closely.

Let f be an additive polynomial in n variables over K, and assume that f is reduced. Let

$$f_i(X_i) = f(0, \ldots, X_i, \ldots, 0)$$

with X_i in the i-th place, and zeros in the other components. By additivity, it follows that

$$f(X_1, \ldots, X_n) = f_1(X_1) + \cdots + f_n(X_n)$$

because the difference of the right-hand side and left-hand side is a reduced polynomial taking the value 0 on $K^{(n)}$. Furthermore, each f_i is an additive polynomial in one variable. We now study such polynomials.

Let $f(X)$ be a reduced polynomial in one variable, which induces a linear map of K into itself. Suppose that there occurs a monomial $a_r X^r$ in f with coefficient $a_r \neq 0$. Then the monomials of degree r in

$$g(X, Y) = f(X + Y) - f(X) - f(Y)$$

are given by

$$a_r(X + Y)^r - a_r X^r - a_r Y^r.$$

We have already seen that g is identically 0. Hence the above expression is identically 0. Hence the polynomial

$$(X + Y)^r - X^r - Y^r$$

is the zero polynomial. It contains the term $rX^{r-1}Y$. Hence if $r > 1$, our field must have characteristic p and r is divisible by p. Write $r = p^m s$ where s is prime to p. Then

$$0 = (X + Y)^r - X^r - Y^r = (X^{p^m} + Y^{p^m})^s - (X^{p^m})^s - (Y^{p^m})^s.$$

Arguing as before, we conclude that $s = 1$.

Hence if f is an additive polynomial in one variable, we have

$$f(X) = \sum_{\nu=0}^{m} a_\nu X^{p^\nu},$$

with $a_\nu \in K$. In characteristic 0, the only additive polynomials in one variable are of type aX with $a \in K$.

As expected, we define $\lambda_1, \ldots, \lambda_n$ to be **algebraically independent** if, whenever f is a reduced polynomial such that $f(\Lambda(x)) = 0$ for all $x \in K$, then f is the zero polynomial.

We shall apply Theorem 12.1 to the case when $\lambda_1, \ldots, \lambda_n$ are automorphisms of a field, and combine Theorem 12.1 with the theorem on the linear independence of characters.

Theorem 12.2. *Let K be an infinite field, and let $\sigma_1, \ldots, \sigma_n$ be the distinct elements of a finite group of automorphisms of K. Then $\sigma_1, \ldots, \sigma_n$ are algebraically independent over K.*

Proof. (Artin). In characteristic 0, Theorem 12.1 and the linear independence of characters show that our assertion is true. Let the characteristic be $p > 0$, and assume that $\sigma_1, \ldots, \sigma_n$ are algebraically dependent.

There exists an additive polynomial $f(X_1, \ldots, X_n)$ in $K[X]$ which is reduced, $f \neq 0$, and such that

$$f(\sigma_1(x), \ldots, \sigma_n(x)) = 0$$

for all $x \in K$. By what we saw above, we can write this relation in the form

$$\sum_{i=1}^{n} \sum_{r=1}^{m} a_{ir}\sigma_i(x)^{p^r} = 0$$

for all $x \in K$, and with not all coefficients a_{ir} equal to 0. Therefore by the linear independence of characters, the automorphisms

$$\{\sigma_i^{p^r}\} \qquad \text{with} \quad i = 1, \ldots, n \quad \text{and} \quad r = 1, \ldots, m$$

cannot be all distinct. Hence we have

$$\sigma_i^{p^r} = \sigma_j^{p^s}$$

with either $i \neq j$ or $r \neq s$. Say $r \leqq s$. For all $x \in K$ we have

$$\sigma_i(x)^{p^r} = \sigma_j(x)^{p^s}.$$

Extracting p-th roots in characteristic p is unique. Hence

$$\sigma_i(x) = \sigma_j(x)^{p^{s-r}} = \sigma_j(x^{p^{s-r}})$$

for all $x \in K$. Let $\sigma = \sigma_j^{-1}\sigma_i$. Then

$$\sigma(x) = x^{p^{s-r}}$$

for all $x \in K$. Taking $\sigma^n = \text{id}$ shows that

$$x = x^{p^{n(s-r)}}$$

for all $x \in K$. Since K is infinite, this can hold only if $s = r$. But in that case, $\sigma_i = \sigma_j$, contradicting the fact that we started with distinct automorphisms.

§13. THE NORMAL BASIS THEOREM

Theorem 13.1. *Let K/k be a finite Galois extension of degree n. Let $\sigma_1, \dots, \sigma_n$ be the elements of the Galois group G. Then there exists an element $w \in K$ such that $\sigma_1 w, \dots, \sigma_n w$ form a basis of K over k.*

Proof. We prove this here only when k is infinite. The case when k is finite can be proved later by methods of linear algebra, as an exercise.

For each $\sigma \in G$, let X_σ be a variable, and let $t_{\sigma,\tau} = X_{\sigma^{-1}\tau}$. Let $X_i = X_{\sigma_i}$. Let

$$f(X_1, \dots, X_n) = \det(t_{\sigma_i, \sigma_j}).$$

Then f is not identically 0, as one sees by substituting 1 for X_{id} and 0 for X_σ if $\sigma \neq \mathrm{id}$. Since k is infinite, f is reduced. Hence the determinant will not be 0 for all $x \in K$ if we substitute $\sigma_i(x)$ for X_i in f. Hence there exists $w \in K$ such that

$$\det(\sigma_i^{-1}\sigma_j(w)) \neq 0.$$

Suppose $a_1, \dots, a_n \in k$ are such that

$$a_1\sigma_1(w) + \cdots + a_n\sigma_n(w) = 0.$$

Apply σ_i^{-1} to this relation for each $i = 1, \dots, n$. Since $a_j \in k$ we get a system of linear equations, regarding the a_j as unknowns. Since the determinant of the coefficients is $\neq 0$, it follows that

$$a_j = 0 \quad \text{for} \quad j = 1, \dots, n$$

and hence that w is the desired element.

Remark. In terms of representations as in Chapters III and XVIII, the normal basis theorem says that the representation of the Galois group on the additive group of the field is the regular representation. One may also say that K is free of dimension 1 over the group ring $k[G]$. Such a result may be viewed as the first step in much more subtle investigations having to do with algebraic number theory. Let K be a number field (finite extension of $\mathbf{Q}$) and let $\mathfrak{o}_K$ be its ring of algebraic integers, which will be defined in Chapter VII, §1. Then one may ask for a description of $\mathfrak{o}_K$ as a $\mathbf{Z}[G]$ module, which is a much more difficult problem. For fundamental work about this problem, see A. Fröhlich, *Galois Module Structures of Algebraic Integers*, Ergebnisse der Math. 3 Folge Vol. 1, Springer Verlag (1983). See also the reference [CCFT 91] given at the end of Chapter III, §1.

§14. INFINITE GALOIS EXTENSIONS

Although we have already given some of the basic theorems of Galois theory already for possibly infinite extensions, the non-finiteness did not really appear in a substantial way. We now want to discuss its role more extensively.

Let K/k be a Galois extension with group G. For each finite Galois subextension F, we have the Galois groups $G_{K/F}$ and $G_{F/k}$. Put $H = G_{K/F}$. Then H has finite index, equal to $\#(G_{F/k}) = [F : k]$. This just comes as a special case of the general Galois theory. We have a canonical homomorphism

$$G \to G/H = G_{F/k}.$$

Therefore by the universal property of the inverse limit, we obtain a homomorphism

$$G \to \varprojlim_{H \in \mathfrak{F}} G/H,$$

where the limit is taken for H in the family $\mathfrak{F}$ of Galois groups $G_{K/F}$ as above.

Theorem 14.1. *The homomorphism $G \to \varprojlim G/H$ is an isomorphism.*

Proof. First the kernel is trivial, because if σ is in the kernel, then σ restricted to every finite subextension of K is trivial, and so is trivial on K. Recall that an element of the inverse limit is a family $\{\sigma_H\}$ with $\sigma_H \in G/H$, satisfying a certain compatibility condition. This compatibility condition means that we may define an element σ of G as follows. Let $\alpha \in K$. Then α is contained in some finite Galois extension $F \subset K$. Let $H = \text{Gal}(K/F)$. Let $\sigma\alpha = \sigma_H\alpha$. The compatibility condition means that $\sigma_H\alpha$ is independent of the choice of F. Then it is immediately verified that σ is an automorphism of K over k, which maps to each σ_H in the canonical map of G into G/H. Hence the map $G \to \varprojlim G/H$ is surjective, thereby proving the theorem.

Remark. For the topological interpretation, see Chapter I, Theorem 10.1, and Exercise 43.

Example. Let $\boldsymbol{\mu}[p^\infty]$ be the union of all groups of roots of unity $\boldsymbol{\mu}[p^n]$, where p is a prime and $n = 1, 2, \ldots$ ranges over the positive integers. Let $K = \mathbf{Q}(\boldsymbol{\mu}[p^\infty])$. Then K is an abelian infinite extension of $\mathbf{Q}$. Let $\mathbf{Z}_p$ be the ring of p-adic integers, and $\mathbf{Z}_p^*$ the group of units. From §3, we know that $(\mathbf{Z}/p^n\mathbf{Z})^*$ is isomorphic to $\text{Gal}(\mathbf{Q}(\boldsymbol{\mu}[p^n]/\mathbf{Q}))$. These isomorphisms are compatible in the tower of p-th roots of unity, so we obtain an isomorphism

$$\mathbf{Z}_p^* \to \text{Gal}(\mathbf{Q}(\boldsymbol{\mu}[p^\infty]/\mathbf{Q})).$$

Towers of cyclotomic fields have been extensively studied by Iwasawa. Cf. a systematic exposition and bibliography in [La 90].

For other types of representations in a group $GL_2(\mathbf{Z}_p)$, see Serre [Se 68], [Se 72], Shimura [Shi 71], and Lang-Trotter [LaT 75]. One general framework in which the representation of Galois groups on roots of unity can be seen has to do with commutative algebraic groups, starting with elliptic curves. Specifically, consider an equation

$$y^2 = 4x^3 - g_2 x - g_3$$

with $g_2, g_3 \in \mathbf{Q}$ and non-zero discriminant: $\Delta = g_2^3 - 27g_3^2 \neq 0$. The set of solutions together with a point at infinity is denoted by E. From complex analysis (or by purely algebraic means), one sees that if K is an extension of $\mathbf{Q}$, then the set of solutions $E(K)$ with $x, y \in K$ and ∞ form a group, called the group of rational points of E in K. One is interested in the torsion group, say $E(\mathbf{Q}^a)_{\text{tor}}$ of points in the algebraic closure, or for a given prime p, in the group $E(\mathbf{Q}^a)[p^r]$ and $E(\mathbf{Q}^a)[p^\infty]$. As an abelian group, there is an isomorphism

$$E(\mathbf{Q}^a)[p^r] \approx (\mathbf{Z}/p^r\mathbf{Z}) \times (\mathbf{Z}/p^r\mathbf{Z}),$$

so the Galois group operates on the points of order p^r via a representation in $GL_2(\mathbf{Z}/p^r\mathbf{Z})$, rather than $GL_1(\mathbf{Z}/p^r\mathbf{Z}) = (\mathbf{Z}/p^r\mathbf{Z})^*$ in the case of roots of unity. Passing to the inverse limit, one obtains a representation of $\text{Gal}(\mathbf{Q}^a/\mathbf{Q}) = G_{\mathbf{Q}}$ in $GL_2(\mathbf{Z}_p)$. One of Serre's theorems is that the image of $G_{\mathbf{Q}}$ in $GL_2(\mathbf{Z}_p)$ is a subgroup of finite index, equal to $GL_2(\mathbf{Z}_p)$ for all but a finite number of primes p, if $\text{End}\, \mathbf{C}\, (E) = \mathbf{Z}$.

More generally, using freely the language of algebraic geometry, when A is a commutative algebraic group, say with coefficients in $\mathbf{Q}$, then one may consider its group of points $A(\mathbf{Q}^a)_{\text{tor}}$, and the representation of $G_{\mathbf{Q}}$ in a similar way. Developing the notions to deal with these situations leads into algebraic geometry.

Instead of considering cyclotomic extensions of a ground field, one may also consider extensions of cyclotomic fields. The following conjecture is due to Shafarevich. See the references at the end of §7.

Conjecture 14.2. *Let $k_0 = \mathbf{Q}(\boldsymbol{\mu})$ be the compositum of all cyclotomic extensions of $\mathbf{Q}$ in a given algebraic closure $\mathbf{Q}^a$. Let k be a finite extension of k_0. Let $G_k = \text{Gal}(\mathbf{Q}^a/k)$. Then G_k is isomorphic to the completion of a free group on countably many generators.*

If G is the free group, then we recall that the completion is the inverse limit $\varprojlim G/H$, taken over all normal subgroups H of finite index. Readers should view this conjecture as being in analogy to the situation with Riemann surfaces, as mentioned in Example 9 of §2. It would be interesting to investigate the extent to which the conjecture remains valid if $\mathbf{Q}(\boldsymbol{\mu})$ is replaced by $\mathbf{Q}(A(\mathbf{Q}^a)_{\text{tor}})$, where A is an elliptic curve. For some results about free groups occurring as Galois groups, see also Wingberg [Wi 91].

Bibliography

[La 90] S. LANG, *Cyclotomic Fields I and II*, Second Edition, Springer Verlag, 1990 (Combined edition from the first editions, 1978, 1980)

[LaT 75] S. LANG and H. TROTTER, *Distribution of Frobenius Elements in GL_2-Extensions of the Rational Numbers*, Springer Lecture Notes **504** (1975)

[Se 68] J.-P. SERRE, *Abelian l-adic Representations and Elliptic Curves*, Benjamin, 1968

[Se 72] J.-P. SERRE, Propriétés galoisiennes des points d'ordre fini des courbes elliptiques, *Invent. Math.* **15** (1972), pp. 259-331

[Shi 71] G. SHIMURA, *Introduction to the arithmetic theory of Automorphic Functions*, Iwanami Shoten and Princeton University Press, 1971

[Wi 91] K. WINGBERG, On Galois groups of p-closed algebraic number fields with restricted ramification, I, *J. reine angew. Math.* **400** (1989), pp. 185–202; and II, *ibid.*, **416** (1991), pp. 187–194

§15. THE MODULAR CONNECTION

This final section gives a major connection between Galois theory and the theory of modular forms, which has arisen since the 1960s.

One fundamental question is whether given a finite group G, there exists a Galois extension K of $\mathbf{Q}$ whose Galois group is G. In Exercise 23 you will prove this when G is abelian.

Already in the nineteenth century, number theorists realized the big difference between abelian and non-abelian extensions, and started understanding abelian extensions. Kronecker stated and gave what are today considered incomplete arguments that every finite abelian extension of $\mathbf{Q}$ is contained in some extension $\mathbf{Q}(\zeta)$, where ζ is a root of unity. The difficulty lay in the peculiarities of the prime 2. The trouble was fixed by Weber at the end of the nineteenth century. Note that the trouble with 2 has been systematic since then. It arose in Artin's conjecture about densities of primitive roots as mentioned in the remarks after Theorem 9.4. It arose in the Grunwald theorem of class field theory (corrected by Wang, cf. Artin-Tate [ArT 68], Chapter 10). It arose in Shafarevich's proof that given a solvable group, there exists a Galois extension of $\mathbf{Q}$ having that group as Galois group, mentioned at the end of §7.

Abelian extensions of a number field F are harder to describe than over the rationals, and the fundamental theory giving a description of such extensions is called class field theory (see the above reference). I shall give one significant example exhibiting the flavor. Let R_F be the ring of algebraic integers in F. It can be shown that R_F is a Dedekind ring. (*Cf.* [La 70], Chapter I, §6, Theorem 2.) Let P be a prime ideal of R_F. Then $P \cap \mathbf{Z} = (p)$ for some prime number p.

Furthermore, R_F/P is a finite field with q elements. Let K be a finite Galois extension of F. It will be shown in Chapter VII that there exists a prime Q of R_K such that $Q \cap R_F = P$. Furthermore, there exists an element

$$\mathrm{Fr}_Q \in G = \mathrm{Gal}(K/F)$$

such that $\mathrm{Fr}_Q(Q) = Q$ and for all $\alpha \in R_K$ we have

$$\mathrm{Fr}_Q \alpha \equiv \alpha^q \bmod Q.$$

We call Fr_Q a **Frobenius element** in the Galois group G associated with Q. (See Chapter VII, Theorem 2.9.) Furthermore, for all but a finite number of Q, two such elements are conjugate to each other in G. We denote any of them by Fr_P. If G is abelian, then there is only one element Fr_P in the Galois group.

Theorem 15.1. *There exists a unique finite abelian extension K of F having the following property. If P_1, P_2 are prime ideals of R_F, then $\mathrm{Fr}_{P_1} = \mathrm{Fr}_{P_2}$ if and only if there is an element α of K such that $\alpha P_1 = P_2$.*

In a similar but more complicated manner, one can characterize all abelian extensions of F. This theory is known as class field theory, developed by Kronecker, Weber, Hilbert, Takagi, and Artin. The main statement concerning the Frobenius automorphism as above is Artin's Reciprocity Law. Artin-Tate's notes give a cohomological account of class field theory. My *Algebraic Number Theory* gives an account following Artin's first proof dating back to 1927, with later simplifications by Artin himself. Both techniques are valuable to know.

Cyclotomic extensions should be viewed in the light of Theorem 15.1. Indeed, let $K = \mathbf{Q}(\zeta)$, where ζ is a primitive n-th root of unity. For a prime $p \nmid n$, we have the Frobenius automorphism Fr_p, whose effect on ζ is $\mathrm{Fr}_p(\zeta) = \zeta^p$. Then

$$\mathrm{Fr}_{p_1} = \mathrm{Fr}_{p_2} \quad \text{if and only if} \quad p_1 \equiv p_2 \bmod n.$$

To encompass both Theorem 15.1 and the cyclotomic case in one framework, one has to formulate the result of class field theory for generalized ideal classes, not just the ordinary ones when two ideals are equivalent if and only if they differ multiplicatively by a non-zero field element. See my *Algebraic Number Theory* for a description of these generalized ideal classes.

The non-abelian case is much more difficult. I shall indicate briefly a special case which gives some of the flavor of what goes on. The problem is to do for non-abelian extensions what Artin did for abelian extensions. Artin went as far as saying that the problem was not to give proofs but to formulate what was to be proved. The insight of Langlands and others in the sixties shows that actually Artin was mistaken. The problem lies in both. Shimura made several computations in this direction involving "modular forms" [Sh 66]. Langlands gave a number of conjectures relating Galois groups with "automorphic forms", which showed that the answer lay in deeper theories, whose formulations, let alone their proofs, were difficult. Great progress was made in the seventies by Serre and Deligne, who proved a first case of Langland's conjecture [DeS 74].

The study of non-abelian Galois groups occurs via their linear "representations". For instance, let l be a prime number. We can ask whether $GL_n(\mathbf{F}_l)$, or $GL_2(\mathbf{F}_l)$, or $PGL_2(\mathbf{F}_l)$ occurs as a Galois group over $\mathbf{Q}$, and "how". The problem is to find natural objects on which the Galois group operates as a linear map, such that we get in a natural way an isomorphism of this Galois group with one of the above linear groups. The theories which indicate in which direction to find such objects are much beyond the level of this course, and lie in the theory of modular functions, involving both analysis and algebra, which form a background for the number theoretic applications. Again I pick a special case to give the flavor.

Let K be a finite Galois extension of $\mathbf{Q}$, with Galois group

$$G = \mathrm{Gal}(K/\mathbf{Q}).$$

Let

$$\rho\colon G \to GL_2(\mathbf{F}_l)$$

be a homomorphism of G into the group of 2×2 matrices over the finite field $\mathbf{F}_l$ for some prime l. Such a homomorphism is called a **representation** of G. From elementary linear algebra, if

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is a 2×2 matrix, we have its trace and determinant defined by

$$\mathrm{tr}(M) = a + d \quad \text{and} \quad \det M = ad - bc.$$

Thus we can take the trace and determinant $\mathrm{tr}\,\rho(\sigma)$ and $\det \rho(\sigma)$ for $\sigma \in G$.

Consider the infinite product with a variable q:

$$\Delta(q) = q \prod_{n=1}^{\infty} (1 - q^n)^{24} = \sum_{n=1}^{\infty} a_n q^n.$$

The coefficients a_n are integers, and $a_1 = 1$.

Theorem 15.2. *For each prime l there exists a unique Galois extension K of $\mathbf{Q}$, with Galois group G, and an injective homomorphism*

$$\rho\colon G \to GL_2(\mathbf{F}_l)$$

having the following property. For all but a finite number of primes p, if a_p is the coefficient of q^p in $\Delta(q)$, then we have

$$\mathrm{tr}\,\rho(\mathrm{Fr}_p) \equiv a_p \bmod l \quad \text{and} \quad \det \rho(\mathrm{Fr}_p) \equiv p^{11} \bmod l.$$

Furthermore, for all primes $l \neq 2, 3, 5, 7, 23, 691$, the image $\rho(G)$ in $GL_2(\mathbf{F}_l)$ consists of those matrices $M \in GL_2(\mathbf{F}_l)$ such that $\det M$ *is an eleventh power in $\mathbf{F}_l^*$.*

The above theorem was conjectured by Serre in 1968 [Se 68]. A proof of the existence as in the first statement was given by Deligne [De 68]. The second statement, describing how big the Galois group actually is in the group of matrices $GL_2(\mathbf{F}_l)$ is due to Serre and Swinnerton-Dyer [Se 72], [SwD 73].

The point of $\Delta(q)$ is that if we put $q = e^{2\pi iz}$, where z is a variable in the upper half-plane, then Δ is a modular form of weight 12. For definitions and an introduction, see the last chapter of [Se 73], [La 73], [La 76], and the following comments. The general result behind Theorem 15.2 for modular forms of weight $\geqq 2$ was given by Deligne [De 73]. For weight 1, it is due to Deligne-Serre [DeS 74]. We summarize the situation as follows.

Let N be a positive integer. To N we associate the subgroups

$$\Gamma(N) \subset \Gamma_1(N) \subset \Gamma_0(N)$$

of $SL_2(\mathbf{Z})$ defined by the conditions for a matrix $\alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbf{Z})$:

$\alpha \in \Gamma(N)$ if and only if $a \equiv d \equiv 1 \bmod N$ and $b \equiv c \equiv 0 \bmod N$;

$\alpha \in \Gamma_1(N)$ if and only if $a \equiv d \equiv 1 \bmod N$ and $c \equiv 0 \bmod N$;

$\alpha \in \Gamma_0(N)$ if and only if $c \equiv 0 \bmod N$.

Let f be a function on the upper half-plane $\mathfrak{H} = \{z \in \mathbf{C}, \operatorname{Im}(z) > 0\}$. Let k be an integer. For

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbf{R}),$$

define $f \circ [\gamma]_k$ (an operation on the right) by

$$f \circ [\gamma]_k(z) = (cz + d)^{-k} f(\gamma z) \quad \text{where} \quad \gamma z = \frac{az + b}{cz + d}.$$

Let Γ be a subgroup of $SL_2(\mathbf{Z})$ containing $\Gamma(N)$. We define f to be **modular of weight k on Γ** if:

$\mathbf{M_k}$ 1. f is holomorphic on $\mathfrak{H}$;

$\mathbf{M_k}$ 2. f is holomorphic at the cusps, meaning that for all $\alpha \in SL_2(\mathbf{Z})$, the function $f \circ [\alpha]_k$ has a power series expansion

$$f \circ [\alpha]_k(z) = \sum_{n=0}^{\infty} a_n e^{2\pi inz/N};$$

$\mathbf{M_k}$ 3. We have $f \circ [\gamma]_k = f$ for all $\gamma \in \Gamma$.

One says that f is **cuspidal** if in **$\mathbf{M_k}$ 2** the power series has a zero; that is, the power starts with $n \geqq 1$.

Suppose that f is modular of weight k on $\Gamma(N)$. Then f is modular on $\Gamma_1(N)$ if and only if $f(z + 1) = f(z)$, or equivalently f has an expansion of the form

$$f(z) = f_\infty(q_z) = \sum_{n=0}^{\infty} a_n q^n \quad \text{where} \quad q = q_z = e^{2\pi i z}.$$

This power series is called the q-**expansion** of f.

Suppose f has weight k on $\Gamma_1(N)$. If $\gamma \in \Gamma_0(N)$ and γ is the above written matrix, then $f \circ [\gamma]_k$ depends only on the image of d in $(\mathbf{Z}/N\mathbf{Z})^*$, and we then denote $f \circ [\gamma]_k$ by $f \circ [d]_k$. Let

$$\varepsilon: (\mathbf{Z}/N\mathbf{Z})^* \to \mathbf{C}^*$$

be a homomorphism (also called a **Dirichlet character**). One says that ε is **odd** if $\varepsilon(-1) = -1$, and **even** if $\varepsilon(-1) = 1$. One says that f is **modular of type** (k, ε) on $\Gamma_0(N)$ if f has weight k on $\Gamma_1(N)$, and

$$f \circ [d]_k = \varepsilon(d) f \quad \text{for all} \quad d \in (\mathbf{Z}/N\mathbf{Z})^*.$$

It is possible to define an algebra of operators on the space of modular forms of given type. This requires more extensive background, and I refer the reader to [La 76] for a systematic exposition. Among all such forms, it is then possible to distinguish some of them which are eigenvectors for this Hecke algebra, or, as one says, eigenfunctions for this algebra. One may then state the Deligne-Serre theorem as follows.

Let $f \neq 0$ be a modular form of type $(1, \varepsilon)$ on $\Gamma_0(N)$, so f has weight 1. Assume that ε is odd. Assume that f is an eigenfunction of the Hecke algebra, with q-expansion $f_\infty = \sum a_n q^n$, normalized so that $a_1 = 1$. Then there exists a unique finite Galois extension K of $\mathbf{Q}$ with Galois group G, and a representation $\rho: G \to GL_2(\mathbf{C})$ (actually an injective homomorphism), such that for all primes $p \nmid N$ the characteristic polynomial of $\rho(\mathrm{Fr}_p)$ is

$$X^2 - a_p X + \varepsilon(p).$$

The representation ρ is irreducible if and only if f is cuspidal.

Note that the representation ρ has values in $GL_2(\mathbf{C})$. For extensive work of Serre and his conjectures concerning representations of Galois groups in $GL_2(\mathbf{F})$ when $\mathbf{F}$ is a finite field, see [Se 87]. Roughly speaking, the general philosophy started by a conjecture of Taniyama-Shimura and the Langlands conjectures is that everything in sight is "modular". Theorem 15.2 and the Deligne-Serre theorem are prototypes of results in this direction. For "modular" representations in $GL_2(\mathbf{F})$, when $\mathbf{F}$ is a finite field, Serre's conjectures have been proved, mostly by Ribet [Ri 90]. As a result, following an idea of Frey, Ribet also showed how the Taniyama-Shimura conjecture implies Fermat's last theorem [Ri 90b]. Note that Serre's conjectures that certain representations in $GL_2(\mathbf{F})$ are modular imply the Taniyama-Shimura conjecture.

Bibliography

[ArT 68] E. ARTIN and J. TATE, *Class Field Theory*, Benjamin-Addison-Wesley, 1968 (reprinted by Addison-Wesley, 1991)

[De 68] P. DELIGNE, Formes modulaires et représentations l-adiques, *Séminaire Bourbaki* 1968–1969, exp. No. 355

[De 73] P. DELIGNE, Formes modulaires et représentations de $GL(2)$, *Springer Lecture Notes* **349** (1973), pp. 55–105

[DeS 74] P. DELIGNE and J. P. SERRE, Formes modulaires de poids 1, *Ann. Sci. ENS* **7** (1974), pp. 507–530

[La 70] S. LANG, *Algebraic Number Theory*, Springer Verlag, reprinted from Addison-Wesley (1970)

[La 73] S. LANG, *Elliptic functions*, Springer Verlag, 1973

[La 76] S. LANG, *Introduction to modular forms*, Springer Verlag, 1976

[Ri 90a] K. RIBET, On modular representations of $\text{Gal}(\overline{Q}/Q)$ arising from modular forms, *Invent. Math.* **100** (1990), pp. 431–476

[Ri 90b] K. RIBET, From the Taniyama-Shimura conjecture to Fermat's last theorem, *Annales de la Fac. des Sci. Toulouse* (1990), pp. 116–139

[Se 68] J.-P. SERRE, Une interprétation des congruences relatives à la fonction de Ramanujan, *Séminaire Delange-Pisot-Poitou*, 1967–1968

[Se 72] J.-P. SERRE, Congruences et formes modulaires (d'après Swinnerton-Dyer), *Séminaire Bourbaki*, 1971–1972

[Se 73] J.-P. SERRE, *A course in arithmetic*, Springer Verlag, 1973

[Se 87] J.-P. SERRE, Sur les représentations modulaires de degré 2 de $\text{Gal}(\overline{Q}/Q)$, *Duke Math. J.* **54** (1987), pp. 179–230

[Shi 66] G. SHIMURA, A reciprocity law in non-solvable extensions, *J. reine angew. Math.* **221** (1966), pp. 209–220

[Shi 71] G. SHIMURA, *Introduction to the arithmetic theory of automorphic functions*, Iwanami Shoten and Princeton University Press, 1971

[SwD 73] H. P. SWINNERTON-DYER, On l-adic representations and congruences for coefficients of modular forms, (Antwerp conference) *Springer Lecture Notes* **350** (1973)

EXERCISES

1. What is the Galois group of the following polynomials?
   (a) $X^3 - X - 1$ over $\mathbf{Q}$.
   (b) $X^3 - 10$ over $\mathbf{Q}$.
   (c) $X^3 - 10$ over $\mathbf{Q}(\sqrt{2})$.
   (d) $X^3 - 10$ over $\mathbf{Q}(\sqrt{-3})$.
   (e) $X^3 - X - 1$ over $\mathbf{Q}(\sqrt{-23})$.
   (f) $X^4 - 5$ over $\mathbf{Q}$, $\mathbf{Q}(\sqrt{5})$, $\mathbf{Q}(\sqrt{-5})$, $\mathbf{Q}(i)$.
   (g) $X^4 - a$ where a is any integer $\neq 0$, $\neq \pm 1$ and is square free. Over $\mathbf{Q}$.

(h) $X^3 - a$ where a is any square-free integer $\geqq 2$. Over $\mathbf{Q}$.
(i) $X^4 + 2$ over $\mathbf{Q}$, $\mathbf{Q}(i)$.
(j) $(X^2 - 2)(X^2 - 3)(X^2 - 5)(X^2 - 7)$ over $\mathbf{Q}$.
(k) Let $p_1, \ldots, p_n$ be distinct prime numbers. What is the Galois group of $(X^2 - p_1) \cdots (X^2 - p_n)$ over $\mathbf{Q}$?
(l) $(X^3 - 2)(X^3 - 3)(X^2 - 2)$ over $\mathbf{Q}(\sqrt{-3})$.
(m) $X^n - t$, where t is transcendental over the complex numbers $\mathbf{C}$ and n is a positive integer. Over $\mathbf{C}(t)$.
(n) $X^4 - t$, where t is as before. Over $\mathbf{R}(t)$.

2. Find the Galois groups over $\mathbf{Q}$ of the following polynomials.
(a) $X^3 + X + 1$ (b) $X^3 - X + 1$ (g) $X^3 + X^2 - 2X - 1$
(c) $X^3 + 2X + 1$ (d) $X^3 - 2X + 1$
(e) $X^3 - X - 1$ (f) $X^3 - 12X + 8$

3. Let $k = \mathbf{C}(t)$ be the field of rational functions in one variable. Find the Galois group over k of the following polynomials:
(a) $X^3 + X + t$ (b) $X^3 - X + t$
(c) $X^3 + tX + 1$ (d) $X^3 - 2tX + t$
(e) $X^3 - X - t$ (f) $X^3 + t^2X - t^3$

4. Let k be a field of characteristic $\neq 2$. Let $c \in k$, $c \notin k^2$. Let $F = k(\sqrt{c})$. Let $\alpha = a + b\sqrt{c}$ with $a, b \in k$ and not both $a, b = 0$. Let $E = F(\sqrt{\alpha})$. Prove that the following conditions are equivalent.
(1) E is Galois over k.
(2) $E = F(\sqrt{\alpha'})$, where $\alpha' = a - b\sqrt{c}$.
(3) Either $\alpha\alpha' = a^2 - cb^2 \in k^2$ or $c\alpha\alpha' \in k^2$.
Show that when these conditions are satisfied, then E is cyclic over k of degree 4 if and only if $c\alpha\alpha' \in k^2$.

5. Let k be a field of characteristic $\neq 2, 3$. Let $f(X)$, $g(X) = X^2 - c$ be irreducible polynomials over k, of degree 3 and 2 respectively. Let D be the discriminant of f. Assume that

$$[k(D^{1/2}) : k] = 2 \quad \text{and} \quad k(D^{1/2}) \neq k(c^{1/2}).$$

Let α be a root of f and β a root of g in an algebraic closure. Prove:
(a) The splitting field of fg over k has degree 12.
(b) Let $\gamma = \alpha + \beta$. Then $[k(\gamma) : k] = 6$.

6. (a) Let K be cyclic over k of degree 4, and of characteristic $\neq 2$. Let $G_{K/k} = \langle\sigma\rangle$. Let E be the unique subfield of K of degree 2 over k. Since $[K : E] = 2$, there exists $\alpha \in K$ such that $\alpha^2 = \gamma \in E$ and $K = E(\alpha)$. Prove that there exists $z \in E$ such that

$$z\sigma z = -1, \quad \sigma\alpha = z\alpha, \quad z^2 = \sigma\gamma/\gamma.$$

(b) Conversely, let E be a quadratic extension of k and let $G_{E/k} = \langle\tau\rangle$. Let $z \in E$ be an element such that $z\tau z = -1$. Prove that there exists $\gamma \in E$ such that $z^2 = \tau\gamma/\gamma$. Then $E = k(\gamma)$. Let $\alpha^2 = \gamma$, and let $K = k(\alpha)$. Show that K is Galois, cyclic of degree 4 over k. Let σ be an extension of τ to K. Show that σ is an automorphism of K which generates $G_{K/k}$, satisfying $\sigma^2\alpha = -\alpha$ and $\sigma\alpha = \pm z\alpha$. Replacing z by $-z$ originally if necessary, one can then have $\sigma\alpha = z\alpha$.

7. (a) Let $K = \mathbf{Q}(\sqrt{a})$ where $a \in \mathbf{Z}$, $a < 0$. Show that K cannot be embedded in a cyclic extension whose degree over $\mathbf{Q}$ is divisible by 4.
   (b) Let $f(X) = X^4 + 30X^2 + 45$. Let α be a root of F. Prove that $\mathbf{Q}(\alpha)$ is cyclic of degree 4 over $\mathbf{Q}$.
   (c) Let $f(X) = X^4 + 4x^2 + 2$. Prove that f is irreducible over $\mathbf{Q}$ and that the Galois group is cyclic.
8. Let $f(X) = X^4 + aX^2 + b$ be an irreducible polynomial over $\mathbf{Q}$, with roots $\pm\alpha$, $\pm\beta$, and splitting field K.
   (a) Show that $\mathrm{Gal}(K/\mathbf{Q})$ is isomorphic to a subgroup of D_8 (the non-abelian group of order 8 other than the quaternion group), and thus is isomorphic to one of the following:
   (i) $\mathbf{Z}/4\mathbf{Z}$ (ii) $\mathbf{Z}/2\mathbf{Z} \times \mathbf{Z}/2\mathbf{Z}$ (iii) D_8.
   (b) Show that the first case happens if and only if
   $$\frac{\alpha}{\beta} - \frac{\beta}{\alpha} \in \mathbf{Q}.$$
   Case (ii) happens if and only if $\alpha\beta \in \mathbf{Q}$ or $\alpha^2 - \beta^2 \in \mathbf{Q}$. Case (iii) happens otherwise. (Actually, in (ii), the case $\alpha^2 - \beta^2 \in \mathbf{Q}$ cannot occur. It corresponds to a subgroup $D_8 \subset S_4$ which is isomorphic to $\mathbf{Z}/2\mathbf{Z} \times \mathbf{Z}/2\mathbf{Z}$, but is not transitive on $\{1, 2, 3, 4\}$).
   (c) Find the splitting field K in $\mathbf{C}$ of the polynomial
   $$X^4 - 4X^2 - 1.$$
   Determine the Galois group of this splitting field over $\mathbf{Q}$, and describe fully the lattices of subfields and of subgroups of the Galois group.
9. Let K be a finite separable extension of a field k, of prime degree p. Let $\theta \in K$ be such that $K = k(\theta)$, and let $\theta_1, \dots, \theta_p$ be the conjugates of θ over k in some algebraic closure. Let $\theta = \theta_1$. If $\theta_2 \in k(\theta)$, show that K is Galois and in fact cyclic over k.
10. Let $f(X) \in \mathbf{Q}[X]$ be a polynomial of degree n, and let K be a splitting field of f over $\mathbf{Q}$. Suppose that $\mathrm{Gal}(K/\mathbf{Q})$ is the symmetric group S_n with $n > 2$.
    (a) Show that f is irreducible over $\mathbf{Q}$.
    (b) If α is a root of f, show that the only automorphism of $\mathbf{Q}(\alpha)$ is the identity.
    (c) If $n \geqq 4$, show that $\alpha^n \notin \mathbf{Q}$.
11. A polynomial $f(X)$ is said to be **reciprocal** if whenever α is a root, then $1/\alpha$ is also a root. We suppose that f has coefficients in a real subfield k of the complex numbers. If f is irreducible over k, and has a nonreal root of absolute value 1, show that f is reciprocal of even degree.
12. What is the Galois group over the rationals of $X^5 - 4X + 2$?
13. What is the Galois group over the rationals of the following polynomials:
    (a) $X^4 + 2X^2 + X + 3$
    (b) $X^4 + 3X^3 - 3X - 2$
    (c) $X^6 + 22X^5 - 9X^4 + 12X^3 - 37X^2 - 29X - 15$
    [*Hint*: Reduce mod 2, 3, 5.]
14. Prove that given a symmetric group S_n, there exists a polynomial $f(X) \in \mathbf{Z}[X]$ with leading coefficient 1 whose Galois group over $\mathbf{Q}$ is S_n. [*Hint*: Reducing mod 2, 3, 5, show that there exists a polynomial whose reductions are such that the Galois group

contains enough cycles to generate S_n. Use the Chinese remainder theorem, also to be able to apply Eisenstein's criterion.]

15. Let K/k be a Galois extension, and let F be an intermediate field between k and K. Let H be the subgroup of $\mathrm{Gal}(K/k)$ mapping F into itself. Show that H is the normalizer of $\mathrm{Gal}(K/F)$ in $\mathrm{Gal}(K/k)$.

16. Let K/k be a finite Galois extension with group G. Let $\alpha \in K$ be such that $\{\sigma\alpha\}_{\sigma\in G}$ is a normal basis. For each subset S of G let $S(\alpha) = \sum_{\sigma\in S}\sigma\alpha$. Let H be a subgroup of G and let F be the fixed field of H. Show that there exists a basis of F over k consisting of elements of the form $S(\alpha)$.

Cyclotomic fields

17. (a) Let k be a field of characteristic $\nmid 2n$, for some odd integer $n \geqq 1$, and let ζ be a primitive n-th root of unity, in k. Show that k also contains a primitive $2n$-th root of unity.
    (b) Let k be a finite extension of the rationals. Show that there is only a finite number of roots of unity in k.

18. (a) Determine which roots of unity lie in the following fields: $\mathbf{Q}(i)$, $\mathbf{Q}(\sqrt{-2})$, $\mathbf{Q}(\sqrt{2})$, $\mathbf{Q}(\sqrt{-3})$, $\mathbf{Q}(\sqrt{3})$, $\mathbf{Q}(\sqrt{-5})$.
    (b) For which integers m does a primitive m-th root of unity have degree 2 over $\mathbf{Q}$?

19. Let ζ be a primitive n-th root of unity. Let $K = \mathbf{Q}(\zeta)$.
    (a) If $n = p^r$ $(r \geqq 1)$ is a prime power, show that $N_{K/\mathbf{Q}}(1 - \zeta) = p$.
    (b) If n is composite (divisible by at least two primes) then $N_{K/\mathbf{Q}}(1 - \zeta) = 1$.

20. Let $f(X) \in \mathbf{Z}[X]$ be a non-constant polynomial with integer coefficients. Show that the values $f(a)$ with $a \in \mathbf{Z}^+$ are divisible by infinitely many primes.

    [*Note*: This is trivial. A much deeper question is whether there are infinitely many a such that $f(a)$ is prime. There are three necessary conditions:

    The leading coefficient of f is positive.
    The polynomial is irreducible.
    The set of values $f(\mathbf{Z}^+)$ has no common divisor > 1.

    A conjecture of Bouniakowski [Bo 1854] states that these conditions are sufficient. The conjecture was rediscovered later and generalized to several polynomials by Schinzel [Sch 58]. A special case is the conjecture that $X^2 + 1$ represents infinitely many primes. For a discussion of the general conjecture and a quantitative version giving a conjectured asymptotic estimate, see Bateman and Horn [BaH 62]. Also see the comments in [HaR 74]. More precisely, let $f_1, \ldots, f_r$ be polynomials with integer coefficients satisfying the first two conditions (positive leading coefficient, irreducible). Let

    $$f = f_1 \cdots f_r$$

    be their product, and assume that f satisfies the third condition. Define:

    $\pi_{(f)}(x)$ = number of positive integers $n \leqq x$ such that $f_1(n), \ldots, f_r(n)$ are all primes.

    (We ignore the finite number of values of n for which some $f_i(n)$ is negative.) The

Bateman-Horn conjecture is that

$$\pi_{(f)}(x) \sim (d_1 \cdots d_r)^{-1}C(f) \int_0^x \frac{1}{(\log t)^r}\, dt,$$

where

$$C(f) = \prod_p \left\{ \left(1 - \frac{1}{p}\right)^{-r} \left(1 - \frac{N_f(p)}{p}\right) \right\},$$

the product being taken over all primes p, and $N_f(p)$ is the number of solutions of the congruence

$$f(n) \equiv 0 \bmod p.$$

Bateman and Horn show that the product converges absolutely. When $r = 1$ and $f(n) = an + b$ with a, b relatively prime integers, $a > 0$, then one gets Dirichlet's theorem that there are infinitely many primes in an arithmetic progression, together with the Dirichlet density of such primes.

[BaH 62] P. T. BATEMAN and R. HORN, A heuristic asymptotic formula concerning the distribution of prime numbers, *Math. Comp.* **16** (1962) pp. 363-367

[Bo 1854] V. BOUNIAKOWSKY, Sur les diviseurs numériques invariables des fonctions rationnelles entières, *Mémoires sc. math. et phys. T. VI* (1854-1855) pp. 307-329

[HaR 74] H. HALBERSTAM and H.-E. RICHERT, *Sieve methods*, Academic Press, 1974

[Sch 58] A. SCHINZEL and W. SIERPINSKI, Sur certaines hypothèses concernant les nombres premiers, *Acta Arith.* **4** (1958) pp. 185-208

21. (a) Let a be a non-zero integer, p a prime, n a positive integer, and $p \nmid n$. Prove that $p \mid \Phi_n(a)$ if and only if a has period n in $(\mathbf{Z}/p\mathbf{Z})^*$.
    (b) Again assume $p \nmid n$ Prove that $p \mid \Phi_n(a)$ for some $a \in \mathbf{Z}$ if and only if $p \equiv 1 \bmod n$. Deduce from this that there are infinitely many primes $\equiv 1 \bmod n$, a special case of Dirichlet's theorem for the existence of primes in an arithmetic progression.

22. Let $F = \mathbf{F}_p$ be the prime field of characteristic p. Let K be the field obtained from F by adjoining all primitive l-th roots of unity, for all prime numbers $l \neq p$. Prove that K is algebraically closed. [*Hint*: Show that if q is a prime number, and r an integer $\geqq 1$, there exists a prime l such that the period of p mod l is q^r, by using the following old trick of Van der Waerden: Let l be a prime dividing the number

$$b = \frac{p^{q^r} - 1}{p^{q^{r-1}} - 1} = (p^{q^{r-1}} - 1)^{q-1} + q(p^{q^{r-1}} - 1)^{q-2} + \cdots + q.$$

If l does not divide $p^{q^{r-1}} - 1$, we are done. Otherwise, $l = q$. But in that case q^2 does not divide b, and hence there exists a prime $l \neq q$ such that l divides b. Then the degree of $F(\zeta_l)$ over F is q^r, so K contains subfields of arbitrary degree over F.]

23. (a) Let G be a finite *abelian* group. Prove that there exists an abelian extension of $\mathbf{Q}$ whose Galois group is G.

(b) Let k be a finite extension of $\mathbf{Q}$, and let G be a finite abelian group. Prove that there exist infinitely many abelian extensions of k whose Galois group is G.

24. Prove that there are infinitely many non-zero integers a, $b \neq 0$ such that $-4a^3 - 27b^2$ is a square in $\mathbf{Z}$.

25. Let k be a field such that every finite extension is cyclic. Show that there exists an automorphism σ of k^a over k such that k is the fixed field of σ.

26. Let $\mathbf{Q}^a$ be a fixed algebraic closure of $\mathbf{Q}$. Let E be a maximal subfield of $\mathbf{Q}^a$ not containing $\sqrt{2}$ (such a subfield exists by Zorn's lemma). Show that every finite extension of E is cyclic. (Your proof should work taking any algebraic irrational number instead of $\sqrt{2}$.)

27. Let k be a field, k^a an algebraic closure, and σ an automorphism of k^a leaving k fixed. Let F be the fixed field of σ. Show that every finite extension of F is cyclic. (The above two problems are examples of Artin, showing how to dig holes in an algebraically closed field.)

28. Let E be an algebraic extension of k such that every non-constant polynomial $f(X)$ in $k[X]$ has at least one root in E. Prove that E is algebraically closed. [*Hint*: Discuss the separable and purely inseparable cases separately, and use the primitive element theorem.]

29. (a) Let K be a cyclic extension of a field F, with Galois group G generated by σ. Assume that the characteristic is p, and that $[K:F] = p^{m-1}$ for some integer $m \geqq 2$. Let β be an element of K such that $\operatorname{Tr}_F^K(\beta) = 1$. Show that there exists an element α in K such that

$$\sigma\alpha - \alpha = \beta^p - \beta.$$

(b) Prove that the polynomial $X^p - X - \alpha$ is irreducible in $K[X]$.

(c) If θ is a root of this polynomial, prove that $F(\theta)$ is a Galois, cyclic extension of degree p^m of F, and that its Galois group is generated by an extension σ^* of σ such that

$$\sigma^*(\theta) = \theta + \beta.$$

30. Let A be an abelian group and let G be a finite cyclic group operating on A [by means of a homomorphism $G \to \operatorname{Aut}(A)$]. Let σ be a generator of G. We define the trace $\operatorname{Tr}_G = \operatorname{Tr}$ on A by $\operatorname{Tr}(x) = \sum_{\tau \in G} \tau x$. Let A_{Tr} denote the kernel of the trace, and let $(1 - \sigma)A$ denote the subgroup of A consisting of all elements of type $y - \sigma y$. Show that $H^1(G, A) \approx A_{\operatorname{Tr}}/(1 - \sigma)A$.

31. Let F be a finite field and K a finite extension of F. Show that the norm N_F^K and the trace Tr_F^K are surjective (as maps from K into F).

32. Let E be a finite separable extension of k, of degree n. Let $W = (w_1, \dots, w_n)$ be elements of E. Let $\sigma_1, \dots, \sigma_n$ be the distinct embeddings of E in k^a over k. Define the **discriminant** of W to be

$$D_{E/k}(W) = \det(\sigma_i w_j)^2.$$

Prove:

(a) If $V = (v_1, \dots, v_n)$ is another set of elements of E and $C = (c_{ij})$ is a matrix of elements of k such that $w_i = \sum c_{ij} v_j$, then

$$D_{E/k}(W) = \det(C)^2 D_{E/k}(V).$$

(b) The discriminant is an element of k.

(c) Let $E = k(\alpha)$ and let $f(X) = \text{Irr}(\alpha, k, X)$. Let $\alpha_1, \ldots, \alpha_n$ be the roots of f and say $\alpha = \alpha_1$. Then

$$f'(\alpha) = \prod_{j=2}^{n} (\alpha - \alpha_j).$$

Show that

$$D_{E/k}(1, \alpha, \ldots, \alpha^{n-1}) = (-1)^{n(n-1)/2} N_k^E(f'(\alpha)).$$

(d) Let the notation be as in (a). Show that $\det(\text{Tr}(w_i w_j)) = (\det(\sigma_i w_j))^2$. [*Hint*: Let A be the matrix $(\sigma_i w_j)$. Show that ${}^t AA$ is the matrix $(\text{Tr}(w_i w_j))$.]

Rational functions

33. Let $K = \mathbf{C}(x)$ where x is transcendental over $\mathbf{C}$, and let ζ be a primitive cube root of unity in $\mathbf{C}$. Let σ be the automorphism of K over $\mathbf{C}$ such that $\sigma x = \zeta x$. Let τ be the automorphism of K over $\mathbf{C}$ such that $\tau x = x^{-1}$. Show that

$$\sigma^3 = 1 = \tau^2 \quad \text{and} \quad \tau\sigma = \sigma^{-1}\tau.$$

Show that the group of automorphisms G generated by σ and τ has order 6 and the subfield F of K fixed by G is the field $\mathbf{C}(y)$ where $y = x^3 + x^{-3}$.

34. Give an example of a field K which is of degree 2 over two distinct subfields E and F respectively, but such that K is not algebraic over $E \cap F$.

35. Let k be a field and X a variable over k. Let

$$\varphi(X) = \frac{f(X)}{g(X)}$$

be a rational function in $k(X)$, expressed as a quotient of two polynomials f, g which are relatively prime. Define the degree of φ to be $\max(\deg f, \deg g)$. Let $Y = \varphi(X)$. (a) Show that the degree of φ is equal to the degree of the field extension $k(X)$ over $k(Y)$ (assuming $Y \notin k$). (b) Show that every automorphism of $k(X)$ over k can be represented by a rational function φ of degree 1, and is therefore induced by a map

$$X \mapsto \frac{aX + b}{cX + d}$$

with $a, b, c, d \in k$ and $ad - bc \neq 0$. (c) Let G be the group of automorphisms of $k(X)$ over k. Show that G is generated by the following automorphisms:

$$\tau_b : X \mapsto X + b, \qquad \sigma_a : X \mapsto aX \quad (a \neq 0), \qquad X \mapsto X^{-1}$$

with $a, b \in k$.

36. Let k be a finite field with q elements. Let $K = k(X)$ be the rational field in one variable. Let G be the group of automorphisms of K obtained by the mappings

$$X \mapsto \frac{aX + b}{cX + d}$$

with a, b, c, d in k and $ad - bc \neq 0$. Prove the following statements:

(a) The order of G is $q^3 - q$.

(b) The fixed field of G is equal to $k(Y)$ where

$$Y = \frac{(X^{q^2} - X)^{q+1}}{(X^q - X)^{q^2+1}}.$$

(c) Let H_1 be the subgroup of G consisting of the mappings $X \mapsto aX + b$ with $a \neq 0$. The fixed field of H_1 is $k(T)$ where $T = (X^q - X)^{q-1}$.

(d) Let H_2 be the subgroup of H_1 consisting of the mappings $X \to X + b$ with $b \in k$. The fixed field of H_2 is equal to $k(Z)$ where $Z = X^q - X$.

Some aspects of Kummer theory

37. Let k be a field of characteristic 0. Assume that for each finite extension E of k, the index $(E^* : E^{*n})$ is finite for every positive integer n. Show that for each positive integer n, there exists only a finite number of abelian extensions of k of degree n.

38. Let $a \neq 0, \neq \pm 1$ be a square-free integer. For each prime number p, let K_p be the splitting field of the polynomial $X^p - a$ over $\mathbf{Q}$. Show that $[K_p : \mathbf{Q}] = p(p-1)$. For each square-free integer $m > 0$, let

$$K_m = \prod_{p|m} K_p$$

be the compositum of all fields K_p for $p|m$. Let $d_m = [K_m : \mathbf{Q}]$ be the degree of K_m over $\mathbf{Q}$. Show that if m is odd then $d_m = \prod_{p|m} d_p$, and if m is even, $m = 2n$ then $d_{2n} = d_n$ or $2d_n$ according as $\sqrt{a}$ is or is not in the field of m-th roots of unity $\mathbf{Q}(\zeta_m)$.

39. Let K be a field of characteristic 0 for simplicity. Let Γ be a finitely generated subgroup of K^*. Let N be an *odd* positive integer. Assume that for each prime $p|N$ we have

$$\Gamma = \Gamma^{1/p} \cap K,$$

and also that $\mathrm{Gal}(K(\boldsymbol{\mu}_N)/K) \approx \mathbf{Z}(N)^*$. Prove the following.

(a) $\Gamma/\Gamma^N \approx \Gamma/(\Gamma \cap K^{*N}) \approx \Gamma K^{*N}/K^{*N}$.

(b) Let $K_N = K(\boldsymbol{\mu}_N)$. Then

$$\Gamma \cap K_N^{*N} = \Gamma^N.$$

[*Hint*: If these two groups are not equal, then for some prime $p|N$ there exists an element $a \in \Gamma$ such that

$$a = b^p \quad \text{with} \quad b \in K_N \quad \text{but} \quad b \notin K.$$

In other words, a is not a p-th power in K but becomes a p-th power in K_N. The equation $x^p - a$ is irreducible over K. Show that b has degree p over $K(\boldsymbol{\mu}_p)$, and that $K(\boldsymbol{\mu}_p, a^{1/p})$ is not abelian over K, so $a^{1/p}$ has degree p over $K(\boldsymbol{\mu}_p)$. Finish the proof yourself.]

(c) Conclude that the natural Kummer map

$$\Gamma/\Gamma^N \to \mathrm{Hom}(H_\Gamma(N), \mu_N)$$

is an isomorphism.

(d) Let $G_\Gamma(N) = \mathrm{Gal}(K(\Gamma^{1/N}, \mu_N)/K)$. Then the commutator subgroup of $G_\Gamma(N)$ is $H_\Gamma(N)$, and in particular $\mathrm{Gal}(K_N/K)$ is the maximal abelian quotient of $G_\Gamma(N)$.

40. Let K be a field and p a prime number not equal to the characteristic of K. Let Γ be a finitely generated subgroup of K^*, and assume that Γ is equal to its own p-division group in K, that is if $z \in K$ and $z^p \in \Gamma$, then $z \in \Gamma$. If p is odd, assume that $\mu_p \subset K$, and if $p = 2$, assume that $\mu_4 \subset K$. Let

$$(\Gamma : \Gamma^p) = p^{r+1}.$$

Show that $\Gamma^{1/p}$ is its own p-division group in $K(\Gamma^{1/p})$, and

$$[K(\Gamma^{1/p^m}) : K] = p^{m(r+1)}$$

for all positive integers m.

41. **Relative invariants (Sato).** Let k be a field and K an extension of k. Let G be a group of automorphisms of K over k, and assume that k is the fixed field of G. (We do not assume that K is algebraic over k.) By a **relative invariant** of G in K we shall mean an element $P \in K$, $P \neq 0$, such that for each $\sigma \in G$ there exists an element $\chi(\sigma) \in k$ for which $P^\sigma = \chi(\sigma)P$. Since σ is an automorphism, we have $\chi(\sigma) \in k^*$. We say that the map $\chi : G \to k^*$ **belongs** to P, and call it a **character**. Prove the following statements:

(a) The map χ above is a homomorphism.

(b) If the same character χ belongs to relative invariants P and Q then there exists $c \in k^*$ such that $P = cQ$.

(c) The relative invariants form a multiplicative group, which we denote by I. Elements $P_1, \ldots, P_m$ of I are called multiplicatively independent mod k^* if their images in the factor group I/k^* are multiplicatively independent, i.e. if given integers $v_1, \ldots, v_m$ such that

$$P_1^{v_1} \cdots P_m^{v_m} = c \in k^*,$$

then $v_1 = \cdots = v_m = 0$.

(d) If $P_1, \ldots, P_m$ are multiplicatively independent mod k^* prove that they are algebraically independent over k. [*Hint*: Use Artin's theorem on characters.]

(e) Assume that $K = k(X_1, \ldots, X_n)$ is the quotient field of the polynomial ring $k[X_1, \ldots, X_n] = k[X]$, and assume that G induces an automorphism of the polynomial ring. Prove: If $F_1(X)$ and $F_2(X)$ are relative invariant polynomials, then their g.c.d. is relative invariant. If $P(X) = F_1(X)/F_2(X)$ is a relative invariant, and is the quotient of two relatively prime polynomials, then $F_1(X)$ and $F_2(X)$ are relative invariants. Prove that the relative invariant polynomials generate I/k^*. Let S be the set of relative invariant polynomials which cannot be factored into a product of two relative invariant polynomials of degrees $\geqq 1$. Show that the elements of S/k^* are multiplicatively independent, and hence that I/k^* is a free abelian group. [If you know about transcendence degree, then using (d) you can conclude that this group is finitely generated.]

42. Let $f(z)$ be a rational function with coefficients in a finite extension of the rationals. Assume that there are infinitely many roots of unity ζ such that $f(\zeta)$ is a root of unity. Show that there exists an integer n such that $f(z) = cz^n$ for some constant c (which is in fact a root of unity).

    This exercise can be generalized as follows: Let Γ_0 be a finitely generated multiplicative group of complex numbers. Let Γ be the group of all complex numbers γ such that γ^m lies in Γ_0 for some integer $m \neq 0$. Let $f(z)$ be a rational function with complex coefficients such that there exist infinitely many $\gamma \in \Gamma$ for which $f(\gamma)$ lies in Γ. Then again, $f(z) = cz^n$ for some c and n. (Cf. *Fundamentals of Diophantine Geometry*.)

43. Let K/k be a Galois extension. We define the **Krull topology** on the group $G(K/k) = G$ by defining a base for open sets to consist of all sets σH where $\sigma \in G$ and $H = G(K/F)$ for some finite extension F of k contained in K.
    (a) Show that if one takes only those sets σH for which F is finite Galois over k then one obtains another base for the same topology.
    (b) The projective limit $\varprojlim G/H$ is embedded in the direct product
    $$\varprojlim_H G/H \to \prod_H G/H.$$
    Give the direct product the product topology. By Tychonoff's theorem in elementary point set topology, the direct product is compact because it is a direct product of finite groups, which are compact (and of course also discrete). Show that the inverse limit $\varprojlim G/H$ is closed in the product, and is therefore compact.
    (c) Conclude that $G(K/k)$ is compact.
    (d) Show that every closed subgroup of finite index in $G(K/k)$ is open.
    (e) Show that the closed subgroups of $G(K/k)$ are precisely those subgroups which are of the form $G(K/F)$ for some extension F of k contained in K.
    (f) Let H be an arbitrary subgroup of G and let F be the fixed field of H. Show that $G(K/F)$ is the closure of H in G.

44. Let k be a field such that every finite extension is cyclic, and having one extension of degree n for each integer n. Show that the Galois group $G = G(k^a/k)$ is the inverse limit $\lim \mathbf{Z}/m\mathbf{Z}$, as $m\mathbf{Z}$ ranges over all ideals of $\mathbf{Z}$, ordered by inclusion. Show that this limit is isomorphic to the direct product of the limits
$$\prod_p \lim_{n\to\infty} \mathbf{Z}/p^n\mathbf{Z} = \prod_p \mathbf{Z}_p$$
taken over all prime numbers p, in other words, it is isomorphic to the product of all p-adic integers.

45. Let k be a perfect field and k^a its algebraic closure. Let $\sigma \in G(k^a/k)$ be an element of infinite order, and suppose k is the fixed field of σ. For each prime p, let K_p be the composite of all cyclic extensions of k of degree a power of p.
    (a) Prove that k^a is the composite of all extensions K_p.
    (b) Prove that either $K_p = k$, or K_p is infinite cyclic over k. In other words, K_p cannot be finite cyclic over k and $\neq k$.
    (c) Suppose $k^a = K_p$ for some prime p, so k^a is an infinite cyclic tower of p-extensions. Let u be a p-adic unit, $u \in \mathbf{Z}_p^*$ such that u does not represent a rational number. Define σ^u, and prove that σ, σ^u are linearly independent

over **Z**, i.e. the group generated by σ and σ^u is free abelian of rank 2. In particular $\{\sigma\}$ and $\{\sigma, \sigma^u\}$ have the same fixed field k.

Witt vectors

46. Let $x_1, x_2, \ldots$ be a sequence of algebraically independent elements over the integers **Z**. For each integer $n \geqq 1$ define

$$x^{(n)} = \sum_{d \mid n} d x_d^{n/d}.$$

Show that x_n can be expressed in terms of $x^{(d)}$ for $d \mid n$, with rational coefficients.

Using vector notation, we call $(x_1, x_2, \ldots)$ the Witt components of the vector x, and call $(x^{(1)}, x^{(2)}, \ldots)$ its **ghost** components. We call x a **Witt vector**.

Define the power series

$$f_x(t) = \prod_{n \geqq 1} (1 - x_n t^n).$$

Show that

$$-t \frac{d}{dt} \log f_x(t) = \sum_{n \geqq 1} x^{(n)} t^n.$$

[By $\frac{d}{dt} \log f(t)$ we mean $f'(t)/f(t)$ if $f(t)$ is a power series, and the derivative $f'(t)$ is taken formally.]

If x, y are two Witt vectors, define their sum and product componentwise *with respect to the ghost components*, i.e.

$$(x \dotplus y)^{(n)} = x^{(n)} \dotplus y^{(n)}.$$

What is $(x + y)_n$? Well, show that

$$f_x(t) f_y(t) = \prod (1 + (x + y)_n t^n) = f_{x+y}(t).$$

Hence $(x + y)_n$ is a polynomial with integer coefficients in $x_1, y_1, \ldots, x_n, y_n$. Also show that

$$f_{xy}(t) = \prod_{d, e \geqq 1} (1 - x_d^{m/d} y_e^{m/e} t^m)^{de/m}$$

where m is the least common multiple of d, e and d, e range over all integers $\geqq 1$. Thus $(xy)_n$ is also a polynomial in $x_1, y_1 \ldots, x_n, y_n$ with integer coefficients. The above arguments are due to Witt (oral communication) and differ from those of his original paper.

If A is a commutative ring, then taking a homomorphic image of the polynomial ring over **Z** into A, we see that we can define addition and multiplication of Witt vectors with components in A, and that these Witt vectors form a ring $W(A)$. Show that W is a functor, i.e. that any ring homomorphism φ of A into a commutative ring A' induces a homomorphism $W(\varphi): W(A) \to W(A')$.

47. Let p be a prime number, and consider the projection of $W(A)$ on vectors whose components are indexed by a power of p. Now use the log to the base p to index these components, so that we write x_n instead of x_{p^n}. For instance, x_0 now denotes what was x_1 previously. For a Witt vector $x = (x_0, x_1, \ldots, x_n, \ldots)$ define

$$Vx = (0, x_0, x_1, \ldots) \quad \text{and} \quad Fx = (x_0^p, x_1^p, \ldots).$$

Thus V is a shifting operator. We have $V \circ F = F \circ V$. Show that

$$(Vx)^{(n)} = px^{(n-1)} \quad \text{and} \quad x^{(n)} = (Fx)^{(n-1)} + p^n x_n.$$

Also from the definition, we have

$$x^{(n)} = x_0^{p^n} + px_1^{p^{n-1}} + \cdots + p^n x_n.$$

48. Let k be a field of characteristic p, and consider $W(k)$. Then V is an additive endomorphism of $W(k)$, and F is a ring homomorphism of $W(k)$ into itself. Furthermore, if $x \in W(k)$ then

$$px = VFx.$$

If $x, y \in W(k)$, then $(V^i x)(V^j y) = V^{i+j}(F^{pj}x \cdot F^{pi}y)$. For $a \in k$ denote by $\{a\}$ the Witt vector $(a, 0, 0, \ldots)$. Then we can write symbolically

$$x = \sum_{i=0}^{\infty} V^i\{x_i\}.$$

Show that if $x \in W(k)$ and $x_0 \neq 0$ then x is a unit in $W(k)$. *Hint*: One has

$$1 - x\{x_0^{-1}\} = Vy$$

and then

$$x\{x_0^{-1}\} \sum_0^{\infty} (Vy)^i = (1 - Vy) \sum_0^{\infty} (Vy)^i = 1.$$

49. Let n be an integer $\geqq 1$ and p a prime number again. Let k be a field of characteristic p. Let $W_n(k)$ be the ring of truncated Witt vectors $(x_0, \ldots, x_{n-1})$ with components in k. We view $W_n(k)$ as an additive group. If $x \in W_n(k)$, define $\wp(x) = Fx - x$. Then $\wp$ is a homomorphism. If K is a Galois extension of k, and $\sigma \in G(K/k)$, and $x \in W_n(K)$ we can define σx to have component $(\sigma x_0, \ldots, \sigma x_{n-1})$. Prove the analogue of Hilbert's Theorem 90 for Witt vectors, and prove that the first cohomology group is trivial. (One takes a vector whose trace is not 0, and finds a coboundary the same way as in the proof of Theorem 10.1).

50. If $x \in W_n(k)$, show that there exists $\xi \in W_n(k^a)$ such that $\wp(\xi) = x$. Do this inductively, solving first for the first component, and then showing that a vector $(0, \alpha_1, \ldots, \alpha_{n-1})$ is in the image of $\wp$ if and only if $(\alpha_1, \ldots, \alpha_{n-1})$ is in the image of $\wp$. Prove inductively that if $\xi, \xi' \in W_n(k')$ for some extension k' of k and if $\wp\xi = \wp\xi'$ then $\xi - \xi'$ is a vector with components in the prime field. Hence the solutions of $\wp\xi = x$ for given $x \in W_n(k)$ all differ by the vectors with components in the prime field, and there are p^n such vectors. We define

$$k(\xi) = k(\xi_0, \ldots, \xi_{n-1}),$$

or symbolically,

$$k(\wp^{-1}x).$$

Prove that it is a Galois extension of k, and show that the cyclic extensions of k, of degree p^n, are precisely those of type $k(\wp^{-1}x)$ with a vector x such that $x_0 \notin \wp k$.

51. Develop the Kummer theory for abelian extensions of k of exponent p^n by using $W_n(k)$. In other words, show that there is a bijection between subgroups B of $W_n(k)$ containing $\wp W_n(k)$ and abelian extensions as above, given by

$$B \mapsto K_B$$

where $K_B = k(\wp^{-1}B)$. All of this is due to Witt, cf. the references at the end of §8, especially [Wi 37]. The proofs are the same, *mutatis mutandis*, as those given for the Kummer theory in the text.

Further Progress and directions

Major progress was made in the 90s concerning some problems mentioned in the chapter. Foremost was Wiles's proof of enough of the Shimura-Taniyama conjecture to imply Fermat's Last Theorem [Wil 95], [TaW 95].

[TaW 95] R. Taylor and A. Wiles, Ring-theoretic properties or certain Hecke algebras, *Annals of Math.* **141** (1995) pp. 553–572

[Wil 95] A. Wiles, Modular elliptic curves and Fermat's last theorem, *Annals. of Math.* **141** (1995) pp. 443–551

Then a proof of the complete Shimura-Taniyama conjecture was given in [BrCDT 01].

[BrCDT 01] C. Breuil, B. Conrad, F. Diamond, R. Taylor, On the modularity of elliptic curves over **Q**: Wild 3-adic exercises, *J. Amer. Math. Soc.* **14** (2001) pp. 843–839

In a quite different direction, Neukirch started the characterization of number fields by their absolute Galois groups [Ne 68], [Ne 69a], [Ne 69b], and proved it for Galois extensions of **Q**. His results were extended and his subsequent conjectures were proved by Ikeda and Uchida [Ik 77], [Uch 77], [Uch 79], [Uch 81]. These results were extended to finitely generated extensions of **Q** (function fields) by Pop [Pop 94], who has a more extensive bibliography on these and related questions of algebraic geometry. For these references, see the bibliography at the end of the book.

CHAPTER VII

Extensions of Rings

It is not always desirable to deal only with field extensions. Sometimes one wants to obtain a field extension by reducing a ring extension modulo a prime ideal. This procedure occurs in several contexts, and so we are led to give the basic theory of Galois automorphisms over rings, looking especially at how the Galois automorphisms operate on prime ideals or the residue class fields. The two examples given after Theorem 2.9 show the importance of working over rings, to get families of extensions in two very different contexts.

Throughout this chapter, A, B, C will denote commutative rings.

§1. INTEGRAL RING EXTENSIONS

In Chapters V and VI we have studied algebraic extensions of fields. For a number of reasons, it is desirable to study algebraic extensions of rings. For instance, given a polynomial with integer coefficients, say $X^5 - X - 1$, one can reduce this polynomial mod p for any prime p, and thus get a polynomial with coefficients in a finite field. As another example, consider the polynomial

$$X^n + s_{n-1}X^{n-1} + \cdots + s_0$$

where $s_{n-1}, \dots, s_0$ are algebraically independent over a field k. This polynomial has coefficients in $k[s_0, \dots, s_{n-1}]$ and by substituting elements of k for $s_0, \dots, s_{n-1}$ one obtains a polynomial with coefficients in k. One can then get

information about polynomials by taking a homomorphism of the ring in which they have their coefficients. This chapter is devoted to a brief description of the basic facts concerning polynomials over rings.

Let M be an A-module. We say that M is **faithful** if, whenever $a \in A$ is such that $aM = 0$, then $a = 0$. We note that A is a faithful module over itself since A contains a unit element. Furthermore, if $A \neq 0$, then a faithful module over A cannot be the 0-module.

Let A be a subring of B. Let $\alpha \in B$. The following conditions are equivalent:

INT 1. The element α is a root of a polynomial

$$X^n + a_{n-1}X^{n-1} + \cdots + a_0$$

with coefficients $a_i \in A$, and degree $n \geqq 1$. (The essential thing here is that the leading coefficient is equal to 1.)

INT 2. The subring $A[\alpha]$ is a finitely generated A-module.

INT 3. There exists a faithful module over $A[\alpha]$ which is a finitely generated A-module.

We prove the equivalence. Assume **INT 1**. Let $g(X)$ be a polynomial in $A[X]$ of degree $\geqq 1$ with leading coefficient 1 such that $g(\alpha) = 0$. If $f(X) \in A[X]$ then

$$f(X) = q(X)g(X) + r(X)$$

with $q, r \in A[X]$ and $\deg r < \deg g$. Hence $f(\alpha) = r(\alpha)$, and we see that if $\deg g = n$, then $1, \alpha, \ldots, \alpha^{n-1}$ are generators of $A[\alpha]$ as a module over A.

An equation $g(X) = 0$ with g as above, such that $g(\alpha) = 0$ is called an **integral equation** for α over A.

Assume **INT 2.** We let the module be $A[\alpha]$ itself.

Assume **INT 3**, and let M be the faithful module over $A[\alpha]$ which is finitely generated over A, say by elements $w_1, \ldots, w_n$. Since $\alpha M \subset M$ there exist elements $a_{ij} \in A$ such that

$$\begin{aligned} \alpha w_1 &= a_{11}w_1 + \cdots + a_{1n}w_n, \\ &\cdots \\ \alpha w_n &= a_{n1}w_1 + \cdots + a_{nn}w_n. \end{aligned}$$

Transposing $\alpha w_1, \ldots, \alpha w_n$ to the right-hand side of these equations, we conclude that the determinant

$$d = \begin{vmatrix} \alpha - a_{11} & & & \\ & \alpha - a_{22} & & -a_{ij} \\ & & \ddots & \\ -a_{ij} & & & \\ & & & \alpha - a_{nn} \end{vmatrix}$$

is such that $dM = 0$. (This will be proved in the chapter when we deal with determinants.) Since M is faithful, we must have $d = 0$. Hence α is a root of the polynomial

$$\det(X\delta_{ij} - a_{ij}),$$

which gives an integral equation for α over A.

An element α satisfying the three conditions **INT 1, 2, 3** is called **integral** over A.

Proposition 1.1. *Let A be an entire ring and K its quotient field. Let α be algebraic over K. Then there exists an element $c \neq 0$ in A such that $c\alpha$ is integral over A.*

Proof. There exists an equation

$$a_n\alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_0 = 0$$

with $a_i \in A$ and $a_n \neq 0$. Multiply it by a_n^{n-1}. Then

$$(a_n\alpha)^n + \cdots + a_0 a_n^{n-1} = 0$$

is an integral equation for $a_n\alpha$ over A. This proves the proposition.

Let $A \subset B$ be subrings of a commutative ring C, and let $\alpha \in C$. If α is integral over A then α is *a fortiori* integral over B. Thus integrality is preserved under lifting. In particular, α is integral over any ring which is intermediate between A and B.

Let B contain A as a subring. We shall say that B is **integral** over A if every element of B is integral over A.

Proposition 1.2. *If B is integral over A and finitely generated as an A-algebra, then B is finitely generated as an A-module.*

Proof. We may prove this by induction on the number of ring generators, and thus we may assume that $B = A[\alpha]$ for some element α integral over A, by considering a tower

$$A \subset A[\alpha_1] \subset A[\alpha_1, \alpha_2] \subset \cdots \subset A[\alpha_1, \ldots, \alpha_n] = B.$$

But we have already seen that our assertion is true in that case, this being part of the definition of integrality.

Just as we did for extension fields, one may define a class $\mathcal{C}$ of extension rings $A \subset B$ to be **distinguished** if it satisfies the analogous properties, namely:

(1) Let $A \subset B \subset C$ be a tower of rings. The extension $A \subset C$ is in $\mathcal{C}$ if and only if $A \subset B$ is in $\mathcal{C}$ and $B \subset C$ is in $\mathcal{C}$.

(2) If $A \subset B$ is in $\mathcal{C}$, if C is any extension ring of A, and if B, C are both subrings of some ring, then $C \subset B[C]$ is in $\mathcal{C}$. (We note that $B[C] = C[B]$ is the smallest ring containing both B and C.)

As with fields, we find formally as a consequence of (1) and (2) that (3) holds, namely:

(3) If $A \subset B$ and $A \subset C$ are in $\mathcal{C}$, and B, C are subrings of some ring, then $A \subset B[C]$ is in $\mathcal{C}$.

Proposition 1.3. *Integral ring extensions form a distinguished class.*

Proof. Let $A \subset B \subset C$ be a tower of rings. If C is integral over A, then it is clear that B is integral over A and C is integral over B. Conversely, assume that each step in the tower is integral. Let $\alpha \in C$. Then α satisfies an integral equation

$$\alpha^n + b_{n-1}\alpha^{n-1} + \cdots + b_0 = 0$$

with $b_i \in B$. Let $B_1 = A[b_0, \ldots, b_{n-1}]$. Then B_1 is a finitely generated A-module by Proposition 1.2, and is obviously faithful. Then $B_1[\alpha]$ is finite over B_1, hence over A, and hence α is integral over A. Hence C is integral over A. Finally let B, C be extension rings of A and assume B integral over A. Assume that B, C are subrings of some ring. Then $C[B]$ is generated by elements of B over C, and each element of B is integral over C. That $C[B]$ is integral over C will follow immediately from our next proposition.

Proposition 1.4. *Let A be a subring of C. Then the elements of C which are integral over A form a subring of C.*

Proof. Let $\alpha, \beta \in C$ be integral over A. Let $M = A[\alpha]$ and $N = A[\beta]$. Then MN contains 1, and is therefore faithful as an A-module. Furthermore, $\alpha M \subset M$ and $\beta N \subset N$. Hence MN is mapped into itself by multiplication with $\alpha \pm \beta$ and $\alpha\beta$. Furthermore MN is finitely generated over A (if $\{w_i\}$ are generators of M and $\{v_j\}$ are generators of N then $\{w_i v_j\}$ are generators of MN). This proves our proposition.

In Proposition 1.4, the set of elements of C which are integral over A is called the **integral closure of A in C**.

Example. Consider the integers $\mathbf{Z}$. Let K be a finite extension of $\mathbf{Q}$. We call K a **number field**. The integral closure of $\mathbf{Z}$ in K is called the ring of **algebraic integers** of K. This is the most classical example.

In algebraic geometry, one considers a finitely generated entire ring R over $\mathbf{Z}$ or over a field k. Let F be the quotient field of R. One then considers the integral closure of R in F, which is proved to be finite over R. If K is a finite extension of F, one also considers the integral closure of R in K.

Proposition 1.5. *Let $A \subset B$ be an extension ring, and let B be integral over A. Let σ be a homomorphism of B. Then $\sigma(B)$ is integral over $\sigma(A)$.*

Proof. Let $\alpha \in B$, and let

$$\alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_0 = 0$$

be an integral equation for α over A. Applying σ yields

$$\sigma(\alpha)^n + \sigma(a_{n-1})\sigma(\alpha)^{n-1} + \cdots + \sigma(a_0) = 0,$$

thereby proving our assertion.

Corollary 1.6. *Let A be an entire ring, k its quotient field, and E a finite extension of k. Let $\alpha \in E$ be integral over A. Then the norm and trace of α (from E to k) are integral over A, and so are the coefficients of the irreducible polynomial satisfied by α over k.*

Proof. For each embedding σ of E over k, $\sigma\alpha$ is integral over A. Since the norm is the product of $\sigma\alpha$ over all such σ (raised to a power of the characteristic), it follows that the norm is integral over A. Similarly for the trace, and similarly for the coefficients of $\mathrm{Irr}(\alpha, k, X)$, which are elementary symmetric functions of the roots.

Let A be an entire ring and k its quotient field. We say that A is **integrally closed** if it is equal to its integral closure in k.

Proposition 1.7. *Let A be entire and factorial. Then A is integrally closed.*

Proof. Suppose that there exists a quotient a/b with $a, b \in A$ which is integral over A, and a prime element p in A which divides b but not a. We have, for some integer $n \geqq 1$, and $a_i \in A$,

$$(a/b)^n + a_{n-1}(a/b)^{n-1} + \cdots + a_0 = 0$$

whence

$$a^n + a_{n-1}ba^{n-1} + \cdots + a_0 b^n = 0.$$

Since p divides b, it must divide a^n, and hence must divide a, contradiction.

Let $f: A \to B$ be a ring-homomorphism (A, B being commutative rings). We recall that such a homomorphism is also called an ***A*-algebra**. We may view B as an A-module. We say that B is integral over A (for this ring-homomorphism f) if B is integral over $f(A)$. This extension of our definition of integrality is useful because there are applications when certain collapsings take place, and we still wish to speak of integrality. Strictly speaking we should not say that B is integral over A, but that f **is an integral ring-homomorphism**, or simply that f is **integral**. We shall use this terminology frequently.

Some of our preceding propositions have immediate consequences for integral ring-homomorphisms; for instance, if $f: A \to B$ and $g: B \to C$ are integral, then $g \circ f: A \to C$ is integral. However, it is not necessarily true that if $g \circ f$ is integral, so is f.

Let $f: A \to B$ be integral, and let S be a multiplicative subset of A. Then we get a homomorphism

$$S^{-1}f: S^{-1}A \to S^{-1}B,$$

where strictly speaking, $S^{-1}B = (f(S))^{-1}B$, and $S^{-1}f$ is defined by

$$(S^{-1}f)(x/s) = f(x)/f(s).$$

It is trivially verified that this is a homomorphism. We have a commutative diagram

$$\begin{array}{ccc} B & \longrightarrow & S^{-1}B \\ {\scriptstyle f}\uparrow & & \uparrow{\scriptstyle S^{-1}f} \\ A & \longrightarrow & S^{-1}A \end{array}$$

the horizontal maps being the canonical ones: $x \to x/1$.

Proposition 1.8. *Let $f: A \to B$ be integral, and let S be a multiplicative subset of A. Then $S^{-1}f: S^{-1}A \to S^{-1}B$ is integral.*

Proof. If $\alpha \in B$ is integral over $f(A)$, then writing $a\beta$ instead of $f(a)\beta$ for $a \in A$ and $\beta \in B$ we have

$$\alpha^n + a_{n-1}\alpha^{n-1} + \cdots + a_0 = 0$$

with $a_i \in A$. Taking the canonical image in $S^{-1}A$ and $S^{-1}B$ respectively, we see that this relation proves the integrality of $\alpha/1$ over $S^{-1}A$, the coefficients being now $a_i/1$.

Proposition 1.9. *Let A be entire and integrally closed. Let S be a multiplicative subset of A, $0 \notin S$. Then $S^{-1}A$ is integrally closed.*

Proof. Let α be an element of the quotient field, integral over $S^{-1}A$. We have an equation

$$\alpha^n + \frac{a_{n-1}}{s_{n-1}}\alpha^{n-1} + \cdots + \frac{a_0}{s_0} = 0,$$

$a_i \in A$ and $s_i \in S$. Let s be the product $s_{n-1} \cdots s_0$. Then it is clear that $s\alpha$ is integral over A, whence in A. Hence α lies in $S^{-1}A$, and $S^{-1}A$ is integrally closed.

Let $\mathfrak{p}$ be a prime ideal of a ring A and let S be the complement of $\mathfrak{p}$ in A. We write $S = A - \mathfrak{p}$. If $f: A \to B$ is an A-algebra (i.e. a ring-homomorphism), we shall write $B_{\mathfrak{p}}$ instead of $S^{-1}B$. We can view $B_{\mathfrak{p}}$ as an $A_{\mathfrak{p}} = S^{-1}A$-module.

Let A be a subring of B. Let $\mathfrak{p}$ be a prime ideal of A and let $\mathfrak{P}$ be a prime ideal of B. We say that $\mathfrak{P}$ **lies above** $\mathfrak{p}$ if $\mathfrak{P} \cap A = \mathfrak{p}$. If that is the case, then the injection $A \to B$ induces an injection of the factor rings

$$A/\mathfrak{p} \to B/\mathfrak{P},$$

and in fact we have a commutative diagram:

$$\begin{array}{ccc} B & \longrightarrow & B/\mathfrak{P} \\ \uparrow & & \uparrow \\ A & \longrightarrow & A/\mathfrak{p} \end{array}$$

the horizontal arrows being the canonical homomorphisms, and the vertical arrows being injections.

If B is integral over A, then $B/\mathfrak{P}$ is integral over $A/\mathfrak{p}$ by Proposition 1.5.

Proposition 1.10. *Let A be a subring of B, let $\mathfrak{p}$ be a prime ideal of A, and assume B integral over A. Then $\mathfrak{p}B \neq B$ and there exists a prime ideal $\mathfrak{P}$ of B lying above $\mathfrak{p}$.*

Proof. We know that $B_\mathfrak{p}$ is integral over $A_\mathfrak{p}$ and that $A_\mathfrak{p}$ is a local ring with maximal ideal $\mathfrak{m}_\mathfrak{p} = S^{-1}\mathfrak{p}$, where $S = A - \mathfrak{p}$. Since we obviously have

$$\mathfrak{p}B_\mathfrak{p} = \mathfrak{p}A_\mathfrak{p}B_\mathfrak{p} = \mathfrak{m}_\mathfrak{p}B_\mathfrak{p},$$

it will suffice to prove our first assertion when A is a local ring. (Note that the existence of a prime ideal $\mathfrak{p}$ implies that $1 \neq 0$, and $\mathfrak{p}B = B$ if and only if $1 \in \mathfrak{p}B$.) In that case, if $\mathfrak{p}B = B$, then 1 has an expression as a finite linear combination of elements of B with coefficients in $\mathfrak{p}$,

$$1 = a_1b_1 + \cdots + a_nb_n$$

with $a_i \in \mathfrak{p}$ and $b_i \in B$. We shall now use notation as if $A_\mathfrak{p} \subset B_\mathfrak{p}$. We leave it to the reader as an exercise to verify that our arguments are valid when we deal only with a canonical homomorphism $A_\mathfrak{p} \to B_\mathfrak{p}$. Let $B_0 = A[b_1, \ldots, b_n]$. Then $\mathfrak{p}B_0 = B_0$ and B_0 is a finite A-module by Proposition 1.2. Hence $B_0 = 0$ by Nakayama's lemma, contradiction. (See Lemma 4.1 of Chapter X.)

To prove our second assertion, note the following commutative diagram:

$$\begin{array}{ccc} B & \longrightarrow & B_\mathfrak{p} \\ \uparrow & & \uparrow \\ A & \longrightarrow & A_\mathfrak{p} \end{array}$$

We have just proved $\mathfrak{m}_\mathfrak{p}B_\mathfrak{p} \neq B_\mathfrak{p}$. Hence $\mathfrak{m}_\mathfrak{p}B_\mathfrak{p}$ is contained in a maximal ideal $\mathfrak{M}$ of $B_\mathfrak{p}$. Taking inverse images, we see that the inverse image of $\mathfrak{M}$ in $A_\mathfrak{p}$ is an ideal containing $\mathfrak{m}_\mathfrak{p}$ (in the case of an inclusion $A_\mathfrak{p} \subset B_\mathfrak{p}$ the inverse image is $\mathfrak{M} \cap A_\mathfrak{p}$). Since $\mathfrak{m}_\mathfrak{p}$ is maximal, we have $\mathfrak{M} \cap A_\mathfrak{p} = \mathfrak{m}_\mathfrak{p}$. Let $\mathfrak{P}$ be the inverse image of $\mathfrak{M}$ in B (in the case of inclusion, $\mathfrak{P} = \mathfrak{M} \cap B$). Then $\mathfrak{P}$ is a prime ideal of B. The inverse image of $\mathfrak{m}_\mathfrak{p}$ in A is simply $\mathfrak{p}$. Taking the inverse image of $\mathfrak{M}$ going around both ways in the diagram, we find that

$$\mathfrak{P} \cap A = \mathfrak{p},$$

as was to be shown.

Proposition 1.11. *Let A be a subring of B, and assume that B is integral over A. Let $\mathfrak{P}$ be a prime ideal of B lying over a prime ideal $\mathfrak{p}$ of A. Then $\mathfrak{P}$ is maximal if and only if $\mathfrak{p}$ is maximal.*

Proof. Assume $\mathfrak{p}$ maximal in A. Then $A/\mathfrak{p}$ is a field, and $B/\mathfrak{P}$ is an entire ring, integral over $A/\mathfrak{p}$. If $\alpha \in B/\mathfrak{P}$, then α is algebraic over $A/\mathfrak{p}$, and we know that $A/\mathfrak{p}[\alpha]$ is a field. Hence every non-zero element of $B/\mathfrak{P}$ is invertible in $B/\mathfrak{P}$, which is therefore a field. Conversely, assume that $\mathfrak{P}$ is maximal in B. Then $B/\mathfrak{P}$ is a field, which is integral over the entire ring $A/\mathfrak{p}$. If $A/\mathfrak{p}$ is not a field, it has a non-zero maximal ideal $\mathfrak{m}$. By Proposition 1.10, there exists a prime ideal $\mathfrak{M}$ of $B/\mathfrak{P}$ lying above $\mathfrak{m}$, $\mathfrak{M} \neq 0$, contradiction.

§2. INTEGRAL GALOIS EXTENSIONS

We shall now investigate the relationship between the Galois theory of a polynomial, and the Galois theory of this same polynomial reduced modulo a prime ideal.

Proposition 2.1. *Let A be an entire ring, integrally closed in its quotient field K. Let L be a finite Galois extension of K with group G. Let $\mathfrak{p}$ be a maximal ideal of A, and let $\mathfrak{P}$, $\mathfrak{Q}$ be prime ideals of the integral closure B of A in L lying above $\mathfrak{p}$. Then there exists $\sigma \in G$ such that $\sigma\mathfrak{P} = \mathfrak{Q}$.*

Proof. Suppose that $\mathfrak{Q} \neq \sigma\mathfrak{P}$ for any $\sigma \in G$. Then $\tau\mathfrak{Q} \neq \sigma\mathfrak{P}$ for any pair of elements $\sigma, \tau \in G$. There exists an element $x \in B$ such that

$$x \equiv 0 \pmod{\sigma\mathfrak{P}}, \qquad \text{all } \sigma \in G$$

$$x \equiv 1 \pmod{\sigma\mathfrak{Q}}, \qquad \text{all } \sigma \in G$$

(use the Chinese remainder theorem). The norm

$$N_K^L(x) = \prod_{\sigma \in G} \sigma x$$

lies in $B \cap K = A$ (because A is integrally closed), and lies in $\mathfrak{P} \cap A = \mathfrak{p}$. But $x \notin \sigma\mathfrak{Q}$ for all $\sigma \in G$, so that $\sigma x \notin \mathfrak{Q}$ for all $\sigma \in G$. This contradicts the fact that the norm of x lies in $\mathfrak{p} = \mathfrak{Q} \cap A$.

If one localizes, one can eliminate the hypothesis that $\mathfrak{p}$ is maximal; just assume that $\mathfrak{p}$ is prime.

Corollary 2.2 *Let A be integrally closed in its quotient field K. Let E be a finite separable extension of K, and B the integral closure of A in E. Let $\mathfrak{p}$ be a maximal ideal of A. Then there exists only a finite number of prime ideals of B lying above $\mathfrak{p}$.*

Proof. Let L be the smallest Galois extension of K containing E. If $\mathfrak{Q}_1$, $\mathfrak{Q}_2$ are two distinct prime ideals of B lying above $\mathfrak{p}$, and $\mathfrak{P}_1$, $\mathfrak{P}_2$ are two prime ideals of the integral closure of A in L lying above $\mathfrak{Q}_1$ and $\mathfrak{Q}_2$ respectively, then $\mathfrak{P}_1 \neq \mathfrak{P}_2$. This argument reduces our assertion to the case that E is Galois over K, and it then becomes an immediate consequence of the proposition.

Let A be integrally closed in its quotient field K, and let B be its integral closure in a finite Galois extension L, with group G. Then $\sigma B = B$ for every $\sigma \in G$. Let $\mathfrak{p}$ be a maximal ideal of A, and $\mathfrak{P}$ a maximal ideal of B lying above $\mathfrak{p}$. We denote by $G_{\mathfrak{P}}$ the subgroup of G consisting of those automorphisms such that $\sigma\mathfrak{P} = \mathfrak{P}$. Then $G_{\mathfrak{P}}$ operates in a natural way on the residue class field $B/\mathfrak{P}$, and leaves $A/\mathfrak{p}$ fixed. To each $\sigma \in G_{\mathfrak{P}}$ we can associate an automorphism $\bar{\sigma}$ of $B/\mathfrak{P}$ over $A/\mathfrak{p}$, and the map given by

$$\sigma \mapsto \bar{\sigma}$$

induces a homomorphism of $G_{\mathfrak{P}}$ into the group of automorphisms of $B/\mathfrak{P}$ over $A/\mathfrak{p}$.

The group $G_{\mathfrak{P}}$ will be called the **decomposition group** of $\mathfrak{P}$. Its fixed field will be denoted by L^{dec}, and will be called the **decomposition field** of $\mathfrak{P}$. Let B^{dec} be the integral closure of A in L^{dec}, and $\mathfrak{Q} = \mathfrak{P} \cap B^{\text{dec}}$. By Proposition 2.1, we know that $\mathfrak{P}$ is the only prime of B lying above $\mathfrak{Q}$.

Let $G = \bigcup \sigma_j G_{\mathfrak{P}}$ be a coset decomposition of $G_{\mathfrak{P}}$ in G. Then the prime ideals $\sigma_j\mathfrak{P}$ are precisely the distinct primes of B lying above $\mathfrak{p}$. Indeed, for two elements $\sigma, \tau \in G$ we have $\sigma\mathfrak{P} = \tau\mathfrak{P}$ if and only if $\tau^{-1}\sigma\mathfrak{P} = \mathfrak{P}$, i.e. $\tau^{-1}\sigma$ lies in $G_{\mathfrak{P}}$. Thus τ, σ lie in the same coset mod $G_{\mathfrak{P}}$.

It is then immediately clear that the decomposition group of a prime $\sigma\mathfrak{P}$ is $\sigma G_{\mathfrak{P}}\sigma^{-1}$.

Proposition 2.3. *The field L^{dec} is the smallest subfield E of L containing K such that $\mathfrak{P}$ is the only prime of B lying above $\mathfrak{P} \cap E$ (which is prime in $B \cap E$).*

Proof. Let E be as above, and let H be the Galois group of L over E. Let $\mathfrak{q} = \mathfrak{P} \cap E$. By Proposition 2.1, all primes of B lying above $\mathfrak{q}$ are conjugate by elements of H. Since there is only one prime, namely $\mathfrak{P}$, it means that H leaves $\mathfrak{P}$ invariant. Hence $G \subset G_{\mathfrak{P}}$ and $E \supset L^{\text{dec}}$. We have already observed that L^{dec} has the required property.

Proposition 2.4. *Notation being as above, we have $A/\mathfrak{p} = B^{\text{dec}}/\mathfrak{Q}$ (under the canonical injection $A/\mathfrak{p} \to B^{\text{dec}}/\mathfrak{Q}$).*

Proof. If σ is an element of G, not in $G_{\mathfrak{P}}$, then $\sigma\mathfrak{P} \neq \mathfrak{P}$ and $\sigma^{-1}\mathfrak{P} \neq \mathfrak{P}$. Let

$$\mathfrak{Q}_\sigma = \sigma^{-1}\mathfrak{P} \cap B^{\text{dec}}.$$

Then $\mathfrak{Q}_\sigma \neq \mathfrak{Q}$. Let x be an element of B^{dec}. There exists an element y of B^{dec} such that

$$y \equiv x \pmod{\mathfrak{Q}}$$

$$y \equiv 1 \pmod{\mathfrak{Q}_\sigma}$$

for each σ in G, but not in $G_{\mathfrak{P}}$. Hence in particular,

$$y \equiv x \pmod{\mathfrak{P}}$$

$$y \equiv 1 \pmod{\sigma^{-1}\mathfrak{P}}$$

for each σ not in $G_{\mathfrak{P}}$. This second congruence yields

$$\sigma y \equiv 1 \pmod{\mathfrak{P}}$$

for all $\sigma \notin G_{\mathfrak{P}}$. The norm of y from L^{dec} to K is a product of y and other factors σy with $\sigma \notin G_{\mathfrak{P}}$. Thus we obtain

$$N_K^{L^{\text{dec}}}(y) \equiv x \pmod{\mathfrak{P}}.$$

But the norm lies in K, and even in A, since it is a product of elements integral over A. This last congruence holds mod $\mathfrak{Q}$, since both x and the norm lie in B^{dec}. This is precisely the meaning of the assertion in our proposition.

If x is an element of B, we shall denote by $\bar{x}$ its image under the homomorphism $B \to B/\mathfrak{P}$. Then $\bar{\sigma}$ is the automorphism of $B/\mathfrak{P}$ satisfying the relation

$$\bar{\sigma}\bar{x} = (\overline{\sigma x}).$$

If $f(X)$ is a polynomial with coefficients in B, we denote by $\bar{f}(X)$ its natural image under the above homomorphism. Thus, if

$$f(X) = b_n X^n + \cdots + b_0,$$

then

$$\bar{f}(X) = \bar{b}_n X^n + \cdots + \bar{b}_0.$$

Proposition 2.5. *Let A be integrally closed in its quotient field K, and let B be its integral closure in a finite Galois extension L of K, with group G. Let $\mathfrak{p}$ be a maximal ideal of A, and $\mathfrak{P}$ a maximal ideal of B lying above $\mathfrak{p}$. Then $B/\mathfrak{P}$ is a normal extension of $A/\mathfrak{p}$, and the map $\sigma \mapsto \bar{\sigma}$ induces a homomorphism of $G_{\mathfrak{P}}$ onto the Galois group of $B/\mathfrak{P}$ over $A/\mathfrak{p}$.*

Proof. Let $\bar{B} = B/\mathfrak{P}$ and $\bar{A} = A/\mathfrak{p}$. Any element of $\bar{B}$ can be written as $\bar{x}$ for some $x \in B$. Let $\bar{x}$ generate a separable subextension of $\bar{B}$ over $\bar{A}$, and let f be the irreducible polynomial for x over K. The coefficients of f lie in A because x is integral over A, and all the roots of f are integral over A. Thus

$$f(X) = \prod_{i=1}^{m} (X - x_i)$$

splits into linear factors in B. Since

$$\bar{f}(X) = \sum_{i=1}^{m} (X - \bar{x}_i)$$

and all the $\bar{x}_i$ lie in $\bar{B}$, it follows that $\bar{f}$ splits into linear factors in $\bar{B}$. We observe that $f(x) = 0$ implies $\bar{f}(\bar{x}) = 0$. Hence $\bar{B}$ is normal over $\bar{A}$, and

$$[\bar{A}(\bar{x}) : \bar{A}] \leqq [K(x) : K] \leqq [L : K].$$

This implies that the maximal separable subextension of $\bar{A}$ in $\bar{B}$ is of finite degree over $\bar{A}$ (using the primitive element theorem of elementary field theory). This degree is in fact bounded by $[L : K]$.

There remains to prove that the map $\sigma \mapsto \bar{\sigma}$ gives a surjective homomorphism of $G_{\mathfrak{P}}$ onto the Galois group of $\bar{B}$ over $\bar{A}$. To do this, we shall give an argument which reduces our problem to the case when $\mathfrak{P}$ is the only prime ideal of B lying above $\mathfrak{p}$. Indeed, by Proposition 2.4, the residue class fields of the ground ring and the ring B^{dec} in the decomposition field are the same. This means that to prove our surjectivity, we may take L^{dec} as ground field. This is the desired reduction, and we can assume $K = L^{\text{dec}}$, $G = G_{\mathfrak{P}}$.

This being the case, take a generator of the maximal separable subextension of $\bar{B}$ over $\bar{A}$, and let it be $\bar{x}$, for some element x in B. Let f be the irreducible polynomial of x over K. Any automorphism of $\bar{B}$ is determined by its effect on $\bar{x}$, and maps $\bar{x}$ on some root of $\bar{f}$. Suppose that $x = x_1$. Given any root x_i of f, there exists an element σ of $G = G_{\mathfrak{P}}$ such that $\sigma x = x_i$. Hence $\bar{\sigma}\bar{x} = \bar{x}_i$. Hence the automorphisms of $\bar{B}$ over $\bar{A}$ induced by elements of G operate transitively on the roots of $\bar{f}$. Hence they give us all automorphisms of the residue class field, as was to be shown.

Corollary 2.6. *Let A be integrally closed in its quotient field K. Let L be a finite Galois extension of K, and B the integral closure of A in L. Let $\mathfrak{p}$ be a maximal ideal of A. Let $\varphi: A \to A/\mathfrak{p}$ be the canonical homomorphism, and let ψ_1, ψ_2 be two homomorphisms of B extending φ in a given algebraic closure of $A/\mathfrak{p}$. Then there exists an automorphism σ of L over K such that*

$$\psi_1 = \psi_2 \circ \sigma.$$

Proof. The kernels of ψ_1, ψ_2 are prime ideals of B which are conjugate by Proposition 2.1. Hence there exists an element τ of the Galois group G such that ψ_1, $\psi_2 \circ \tau$ have the same kernel. Without loss of generality, we may therefore assume that ψ_1, ψ_2 have the same kernel $\mathfrak{P}$. Hence there exists an automorphism ω of $\psi_1(B)$ onto $\psi_2(B)$ such that $\omega \circ \psi_1 = \psi_2$. There exists an element σ of $G_{\mathfrak{P}}$ such that $\omega \circ \psi_1 = \psi_1 \circ \sigma$, by the preceding proposition. This proves what we wanted.

Remark. In all the above propositions, we could assume $\mathfrak{p}$ prime instead of maximal. In that case, one has to localize at $\mathfrak{p}$ to be able to apply our proofs.

In the above discussions, the kernel of the map

$$G_{\mathfrak{P}} \to \bar{G}_{\mathfrak{P}}$$

is called the **inertia group** of $\mathfrak{P}$. It consists of those automorphisms of $G_{\mathfrak{P}}$ which induce the trivial automorphism on the residue class field. Its fixed field is called the **inertia field**, and is denoted by L^{in}.

Corollary 2.7. *Let the assumptions be as in Corollary* 2.6 *and assume that $\mathfrak{P}$ is the only prime of B lying above $\mathfrak{p}$. Let $f(X)$ be a polynomial in $A[X]$ with leading coefficient* 1. *Assume that f is irreducible in $K[X]$, and has a root α in B. Then the reduced polynomial $\bar{f}$ is a power of an irreducible polynomial in $\bar{A}[X]$.*

Proof. By Corollary 2.6, we know that any two roots of $\bar{f}$ are conjugate under some isomorphism of $\bar{B}$ over $\bar{A}$, and hence that $\bar{f}$ cannot split into relative prime polynomials. Therefore, $\bar{f}$ is a power of an irreducible polynomial.

Proposition 2.8. *Let A be an entire ring, integrally closed in its quotient field K. Let L be a finite Galois extension of K. Let $L = K(\alpha)$, where α is integral over A, and let*

$$f(X) = X^n + a_{n-1}X^{n-1} + \cdots + a_0$$

be the irreducible polynomial of α over k, with $a_i \in A$. Let $\mathfrak{p}$ be a maximal ideal in A, let $\mathfrak{P}$ be a prime ideal of the integral closure B of A in L, $\mathfrak{P}$ lying above $\mathfrak{p}$. Let $\bar{f}(X)$ be the reduced polynomial with coefficients in $A/\mathfrak{p}$. Let $G_{\mathfrak{P}}$ be the decomposition group. If $\bar{f}$ has no multiple roots, then the map $\sigma \mapsto \bar{\sigma}$ has trivial kernel, and is an isomorphism of $G_{\mathfrak{P}}$ on the Galois group of $\bar{f}$ over $A/\mathfrak{p}$.

Proof. Let

$$f(X) = \prod (X - x_i)$$

be the factorization of f in L. We know that all $x_i \in B$. If $\sigma \in G_{\mathfrak{P}}$, then we denote by $\bar{\sigma}$ the homomorphic image of σ in the group $\bar{G}_{\mathfrak{P}}$, as before. We have

$$\bar{f}(X) = \prod (X - \bar{x}_i).$$

Suppose that $\bar{\sigma}\bar{x}_i = \bar{x}_i$ for all i. Since $(\overline{\sigma x_i}) = \bar{\sigma}\bar{x}_i$, and since $\bar{f}$ has no multiple roots, it follows that σ is also the identity. Hence our map is injective, the inertia group is trivial. The field $\bar{A}[\bar{x}_1, \ldots, \bar{x}_n]$ is a subfield of $\bar{B}$ and any auto-

morphism of $\bar{B}$ over $\bar{A}$ which restricts to the identity on this subfield must be the identity, because the map $G_{\mathfrak{P}} \to \bar{G}_{\mathfrak{P}}$ is onto the Galois group of $\bar{B}$ over $\bar{A}$. Hence $\bar{B}$ is purely inseparable over $\bar{A}[\bar{x}_1, \dots, \bar{x}_n]$ and therefore $G_{\mathfrak{P}}$ is isomorphic to the Galois group of $\bar{f}$ over $\bar{A}$.

Proposition 2.8 is only a special case of the more-general situation when the root of a polynomial does not necessarily generate a Galois extension. We state a version useful to compute Galois groups.

Theorem 2.9. *Let A be an entire ring, integrally closed in its quotient field K. Let $f(X) \in A[X]$ have leading coefficient* 1 *and be irreducible over K (or A, it's the same thing). Let $\mathfrak{p}$ be a maximal ideal of A and let $\bar{f} = f$ mod $\mathfrak{p}$. Suppose that $\bar{f}$ has no multiple roots in an algebraic closure of $A/\mathfrak{p}$. Let L be a splitting field for f over K, and let B be the integral closure of A in L. Let $\mathfrak{P}$ be any prime of B above $\mathfrak{p}$ and let a bar denote reduction mod $\mathfrak{p}$. Then the map*

$$G_{\mathfrak{P}} \to \bar{G}_{\mathfrak{P}}$$

is an isomorphism of $G_{\mathfrak{P}}$ with the Galois group of $\bar{f}$ over $\bar{A}$.

Proof. Let $(\alpha_1, \dots, \alpha_n)$ be the roots of f in B and let $(\bar{\alpha}_1, \dots, \bar{\alpha}_n)$ be their reductions mod $\mathfrak{P}$. Since

$$f(X) = \prod_{i=1}^{n} (X - \alpha_i),$$

it follows that

$$\bar{f}(X) = \prod_{i=1}^{n} (X - \bar{\alpha}_i).$$

Any element of G is determined by its effect as a permutation of the roots, and for $\sigma \in G_{\mathfrak{P}}$, we have

$$\bar{\sigma}\,\bar{\alpha}_i = \overline{\sigma\alpha}_i.$$

Hence if $\bar{\sigma} = \text{id}$ then $\sigma = \text{id}$, so the map $G_{\mathfrak{P}} \to \bar{G}_{\mathfrak{P}}$ is injective. It is surjective by Proposition 2.5, so the theorem is proved.

This theorem justifies the statement used to compute Galois groups in Chapter VI, §2.

Theorem 2.9 gives a very efficient tool for analyzing polynomials over a ring.

Example. Consider the "generic" polynomial

$$f_w(X) = X^n + w_{n-1}X^{n-1} + \dots + w_0$$

where $w_0, \ldots, w_{n-1}$ are algebraically independent over a field k. We know that the Galois group of this polynomial over the field $K = k(w_0, \ldots, w_{n-1})$ is the symmetric group. Let $t_1, \ldots, t_n$ be the roots. Let α be a generator of the splitting field L; that is, $L = K(\alpha)$. Without loss of generality, we can select α to be integral over the ring $k[w_0, \ldots, w_{n-1}]$(multiply any given generator by a suitably chosen polynomial and use Proposition 1.1). Let $g_w(X)$ be the irreducible polynomial of α over $k(w_0, \ldots, w_{n-1})$. The coefficients of g are polynomials in (w). If we can substitute values (a) for (w) with $a_0, \ldots, a_{n-1} \in k$ such that g_a remains irreducible, then by Proposition 2.8 we conclude at once that the Galois group of g_a is the symmetric group also. Similarly, if a finite Galois extension of $k(w_0, \ldots, w_{n-1})$ has Galois group G, then we can do a similar substitution to get a Galois extension of k having Galois group G, provided the special polynomial g_a remains irreducible.

Example. Let K be a number field; that is, a finite extension of $\mathbf{Q}$. Let $\mathfrak{o}$ be the ring of algebraic integers. Let L be a finite Galois extension of K and $\mathfrak{O}$ the algebraic integers in L. Let $\mathfrak{p}$ be a prime of $\mathfrak{o}$ and $\mathfrak{P}$ a prime of $\mathfrak{O}$ lying above $\mathfrak{p}$. Then $\mathfrak{o}/\mathfrak{p}$ is a finite field, say with q elements. Then $\mathfrak{O}/\mathfrak{P}$ is a finite extension of $\mathfrak{o}/\mathfrak{p}$, and by the theory of finite fields, there is a unique element in $\bar{G}_{\mathfrak{P}}$, called the **Frobenius element** $\overline{\mathrm{Fr}}_{\mathfrak{P}}$, such that $\overline{\mathrm{Fr}}_{\mathfrak{P}}(\bar{x}) = \bar{x}^q$ for $\bar{x} \in \mathfrak{O}/\mathfrak{P}$. The conditions of Theorem 2.9 are satisfied for all but a finite number of primes $\mathfrak{p}$, and for such primes, there is a unique element $\mathrm{Fr}_{\mathfrak{P}} \in G_{\mathfrak{P}}$ such that $\mathrm{Fr}_{\mathfrak{P}}(x) \equiv x^q \bmod \mathfrak{P}$ for all $x \in \mathfrak{O}$. We call $\mathrm{Fr}_{\mathfrak{P}}$ the **Frobenius element** in $G_{\mathfrak{P}}$. Cf. Chapter VI, §15, where some of the significance of the Frobenius element is explained.

§3. EXTENSION OF HOMOMORPHISMS

When we first discussed the process of localization, we considered very briefly the extension of a homomorphism to a local ring. In our discussion of field theory, we also described an extension theorem for embeddings of one field into another. We shall now treat the extension question in full generality.

First we recall the case of a local ring. Let A be a commutative ring and $\mathfrak{p}$ a prime ideal. We know that the local ring $A_{\mathfrak{p}}$ is the set of all fractions x/y, with $x, y \in A$ and $y \notin \mathfrak{p}$. Its maximal ideal consists of those fractions with $x \in \mathfrak{p}$. Let L be a field and let $\varphi\colon A \to L$ be a homomorphism whose kernel is $\mathfrak{p}$. Then we can extend φ to a homomorphism of $A_{\mathfrak{p}}$ into L by letting

$$\varphi(x/y) = \varphi(x)/\varphi(y)$$

if x/y is an element of $A_{\mathfrak{p}}$ as above.

Second, we have integral ring extensions. Let $\mathfrak{o}$ be a local ring with maximal ideal $\mathfrak{m}$, let B be integral over $\mathfrak{o}$, and let $\varphi : \mathfrak{o} \to L$ be a homomorphism of $\mathfrak{o}$

into an algebraically closed field L. We assume that the kernel of φ is $\mathfrak{m}$. By Proposition 1.10, we know that there exists a maximal ideal $\mathfrak{M}$ of B lying above $\mathfrak{m}$, i.e. such that $\mathfrak{M} \cap \mathfrak{o} = \mathfrak{m}$. Then $B/\mathfrak{M}$ is a field, which is an algebraic extension of $\mathfrak{o}/\mathfrak{m}$, and $\mathfrak{o}/\mathfrak{m}$ is isomorphic to the subfield $\varphi(\mathfrak{o})$ of L because the kernel of φ is $\mathfrak{m}$.

We can find an isomorphism of $\mathfrak{o}/\mathfrak{m}$ onto $\varphi(\mathfrak{o})$ such that the composite homomorphism

$$\mathfrak{o} \to \mathfrak{o}/\mathfrak{m} \to L$$

is equal to φ. We now embed $B/\mathfrak{M}$ into L so as to make the following diagram commutative:

$$\begin{array}{ccccc} B & \longrightarrow & B/\mathfrak{M} & & \\ \uparrow & & \uparrow & \searrow & \\ \mathfrak{o} & \longrightarrow & \mathfrak{o}/\mathfrak{m} & \longrightarrow & L \end{array}$$

and in this way get a homomorphism of B into L which extends φ.

Proposition 3.1. *Let A be a subring of B and assume that B is integral over A. Let $\varphi : A \to L$ be a homomorphism into a field L which is algebraically closed. Then φ has an extension to a homomorphism of B into L.*

Proof. Let $\mathfrak{p}$ be the kernel of φ and let S be the complement of $\mathfrak{p}$ in A. Then we have a commutative diagram

$$\begin{array}{ccc} B & \longrightarrow & S^{-1}B \\ \uparrow & & \uparrow \\ A & \longrightarrow & S^{-1}A = A_{\mathfrak{p}} \end{array}$$

and φ can be factored through the canonical homomorphism of A into $S^{-1}A$. Furthermore, $S^{-1}B$ is integral over $S^{-1}A$. This reduces the question to the case when we deal with a local ring, which has just been discussed above.

Theorem 3.2. *Let A be a subring of a field K and let $x \in K$, $x \neq 0$. Let $\varphi : A \to L$ be a homomorphism of A into an algebraically closed field L. Then φ has an extension to a homomorphism of $A[x]$ or $A[x^{-1}]$ into L.*

Proof. We may first extend φ to a homomorphism of the local ring $A_{\mathfrak{p}}$, where $\mathfrak{p}$ is the kernel of φ. Thus without loss of generality, we may assume that A is a local ring with maximal ideal $\mathfrak{m}$. Suppose that

$$\mathfrak{m}A[x^{-1}] = A[x^{-1}].$$

Then we can write

$$1 = a_0 + a_1 x^{-1} + \cdots + a_n x^{-n}$$

with $a_i \in \mathfrak{m}$. Multiplying by x^n we obtain

$$(1 - a_0)x^n + b_{n-1}x^{n-1} + \cdots + b_0 = 0$$

with suitable elements $b_i \in A$. Since $a_0 \in \mathfrak{m}$, it follows that $1 - a_0 \notin \mathfrak{m}$ and hence $1 - a_0$ is a unit in A because A is assumed to be a local ring. Dividing by $1 - a_0$ we see that x is integral over A, and hence that our homomorphism has an extension to $A[x]$ by Proposition 3.1.

If on the other hand we have

$$\mathfrak{m}A[x^{-1}] \neq A[x^{-1}]$$

then $\mathfrak{m}A[x^{-1}]$ is contained in some maximal ideal $\mathfrak{P}$ of $A[x^{-1}]$ and $\mathfrak{P} \cap A$ contains $\mathfrak{m}$. Since $\mathfrak{m}$ is maximal, we must have $\mathfrak{P} \cap A = \mathfrak{m}$. Since φ and the canonical map $A \to A/\mathfrak{m}$ have the same kernel, namely $\mathfrak{m}$, we can find an embedding ψ of $A/\mathfrak{m}$ into L such that the composite map

$$A \to A/\mathfrak{m} \xrightarrow{\psi} L$$

is equal to φ. We note that $A/\mathfrak{m}$ is canonically embedded in $B/\mathfrak{P}$ where $B = A[x^{-1}]$, and extend ψ to a homomorphism of $B/\mathfrak{P}$ into L, which we can do whether the image of x^{-1} in $B/\mathfrak{P}$ is transcendental or algebraic over $A/\mathfrak{m}$. The composite $B \to B/\mathfrak{P} \to L$ gives us what we want.

Corollary 3.3. *Let A be a subring of a field K and let L be an algebraically closed field. Let $\varphi : A \to L$ be a homomorphism. Let B be a maximal subring of K to which φ has an extension homomorphism into L. Then B is a local ring and if $x \in K$, $x \neq 0$, then $x \in B$ or $x^{-1} \in B$.*

Proof. Let S be the set of pairs (C, ψ) where C is a subring of K and $\psi : C \to L$ is a homomorphism extending φ. Then S is not empty (containing (A, φ)], and is partially ordered by ascending inclusion and restriction. In other words, $(C, \psi) \leqq (C', \psi')$ if $C \subset C'$ and the restriction of ψ' to C is equal to ψ. It is clear that S is inductively ordered, and by Zorn's lemma there exists a maximal element, say (B, ψ_0). Then first B is a local ring, otherwise ψ_0 extends to the local ring arising from the kernel, and second, B has the desired property according to Theorem 3.2.

Let B be a subring of a field K having the property that given $x \in K$, $x \neq 0$, then $x \in B$ or $x^{-1} \in B$. Then we call B a **valuation ring** in K. We shall study such rings in greater detail in Chapter XII. However, we shall also give some applications in the next chapter, so we make some more comments here.

Let F be a field. We let the symbol ∞ satisfy the usual algebraic rules. If $a \in F$, we define

$$a \pm \infty = \infty, \qquad a \cdot \infty = \infty \quad \text{if} \quad a \neq 0,$$

$$\infty \cdot \infty = \infty, \qquad \frac{1}{0} = \infty \quad \text{and} \quad \frac{1}{\infty} = 0.$$

The expressions $\infty \pm \infty$, $0 \cdot \infty$, $0/0$, and ∞/∞ are not defined.

A **place** φ of a field K into a field F is a mapping

$$\varphi : K \to \{F, \infty\}$$

of K into the set consisting of F and ∞ satisfying the usual rules for a homomorphism, namely

$$\varphi(a + b) = \varphi(a) + \varphi(b),$$

$$\varphi(ab) = \varphi(a)\varphi(b)$$

whenever the expressions on the right-hand side of these formulas are defined, and such that $\varphi(1) = 1$. We shall also say that the place is ***F*-valued**. The elements of K which are not mapped into ∞ will be called **finite** under the place, and the others will be called **infinite**.

The reader will verify at once that the set $\mathfrak{o}$ of elements of K which are finite under a place is a valuation ring of K. The maximal ideal consists of those elements x such that $\varphi(x) = 0$. Conversely, if $\mathfrak{o}$ is a valuation ring of K with maximal ideal $\mathfrak{m}$, we let $\varphi : \mathfrak{o} \to \mathfrak{o}/\mathfrak{m}$ be the canonical homomorphism, and define $\varphi(x) = \infty$ for $x \in K$, $x \notin \mathfrak{o}$. Then it is trivially verified that φ is a place.

If $\varphi_1 : K \to \{F_1, \infty\}$ and $\varphi_2 : K \to \{F_2, \infty\}$ are places of K, we take their restrictions to their images. We may therefore assume that they are surjective. We shall say that they are **equivalent** if there exists an isomorphism $\lambda : F_1 \to F_2$ such that $\varphi_2 = \varphi_1 \circ \lambda$. (We put $\lambda(\infty) = \infty$.) One sees that two places are equivalent if and only if they have the same valuation ring. It is clear that there is a bijection between equivalence classes of places of K, and valuation rings of K. A place is called trivial if it is injective. The valuation ring of the trivial place is simply K itself.

As with homomorphisms, we observe that the composite of two places is also a place (trivial verification).

It is often convenient to deal with places instead of valuation rings, just as it is convenient to deal with homomorphisms and not always with canonical homomorphisms or a ring modulo an ideal.

The general theory of valuations and valuation rings is due to Krull, Allgemeine Bewertungstheorie, *J. reine angew. Math.* **167** (1932), pp. 169-196. However, the extension theory of homomorphisms as above was realized only around 1945 by Chevalley and Zariski.

We shall now give some examples of places and valuation rings.

Example 1. Let p be a prime number. Let $\mathbf{Z}_{(p)}$ be the ring of all rational numbers whose denominator is not divisible by p. Then $\mathbf{Z}_{(p)}$ is a valuation ring. The maximal ideal consists of those rational numbers whose numerator is divisible by p.

Example 2. Let k be a field and $R = k[X]$ the polynomial ring in one variable. Let $p = p(X)$ be an irreducible polynomial. Let $\mathfrak{o}$ be the ring of rational functions whose denominator is not divisible by p. Then $\mathfrak{o}$ is a valuation ring, similar to that of Example 1.

Example 3. Let R be the ring of power series $k[[X]]$ in one variable. Then R is a valuation ring, whose maximal ideal consists of those power series divisible by X. The residue class field is k itself.

Example 4. Let $R = k[[X_1, \ldots, X_n]]$ be the ring of power series in several variables. Then R is not a valuation ring, but R is imbedded in the field of repeated power series $k((X_1))((X_2)) \cdots ((X_n)) = K_n$. By Example 3, there is a place of K_n which is K_{n-1}-valued. By induction and composition, we can define a k-valued place of K_n. Since the field of rational functions $k(X_1, \ldots, X_n)$ is contained in K_n, the restriction of this place to $k(X_1, \ldots, X_n)$ gives a k-valued place of the field of rational functions in n variables.

Example 5. In Chapter XI we shall consider the notion of ordered field. Let k be an ordered subfield of an ordered field K. Let $\mathfrak{o}$ be the subset of elements of K which are not infinitely large with respect to k. Let $\mathfrak{m}$ be the subset of elements of $\mathfrak{o}$ which are infinitely small with respect to k. Then $\mathfrak{o}$ is a valuation ring in K and $\mathfrak{m}$ is its maximal ideal.

The following property of places will be used in connection with projective space in the next chapter.

Proposition 3.4. *Let $\varphi: K \to \{L, \infty\}$ be an L-valued place of K. Given a finite number of non-zero elements $x_1, \ldots, x_n \in K$ there exists an index j such that φ is finite on x_i/x_j for $i = 1, \ldots, n$.*

Proof. Let B be the valuation ring of the place. Define $x_i \leqq x_j$ to mean that $x_i/x_j \in B$. Then the relation $\leqq$ is transitive, that is if $x_i \leqq x_j$ and $x_j \leqq x_r$ then $x_i \leqq x_r$. Furthermore, by the property of a valuation ring, we always have $x_i \leqq x_j$ or $x_j \leqq x_i$ for all pairs of indices i, j. Hence we may order our elements, and we select the index j such that $x_i \leqq x_j$ for all i. This index j satisfies the requirement of the proposition.

We can obtain a characterization of integral elements by means of valuation rings. We shall use the following terminology. If $\mathfrak{o}$, $\mathfrak{O}$ are local rings with maximal ideals $\mathfrak{m}$, $\mathfrak{M}$ respectively, we shall say that $\mathfrak{O}$ **lies above** $\mathfrak{o}$ if $\mathfrak{o} \subset \mathfrak{O}$ and $\mathfrak{M} \cap \mathfrak{o} = \mathfrak{m}$. We then have a canonical injection $\mathfrak{o}/\mathfrak{m} \to \mathfrak{O}/\mathfrak{M}$.

Proposition 3.5. *Let $\mathfrak{o}$ be a local ring contained in a field L. An element x of L is integral over $\mathfrak{o}$ if and only if x lies in every valuation ring $\mathfrak{O}$ of L lying above $\mathfrak{o}$.*

Proof. Assume that x is not integral over $\mathfrak{o}$. Let $\mathfrak{m}$ be the maximal ideal of $\mathfrak{o}$. Then the ideal $(\mathfrak{m}, 1/x)$ of $\mathfrak{o}[1/x]$ cannot be the entire ring, otherwise we can write

$$-1 = a_n(1/x)^n + \cdots + a_1(1/x) + y$$

with $y \in \mathfrak{m}$ and $a_i \in \mathfrak{o}$. From this we get

$$(1 + y)x^n + \cdots + a_n = 0.$$

But $1 + y$ is not in $\mathfrak{m}$, hence is a unit of $\mathfrak{o}$. We divide the equation by $1 + y$ to conclude that x is integral over $\mathfrak{o}$, contrary to our hypothesis. Thus $(\mathfrak{m}, 1/x)$ is not the entire ring, and is contained in a maximal ideal $\mathfrak{P}$, whose intersection with $\mathfrak{o}$ contains $\mathfrak{m}$ and hence must be equal to $\mathfrak{m}$. Extending the canonical homomorphism $\mathfrak{o}[1/x] \to \mathfrak{o}[1/x]/\mathfrak{P}$ to a homomorphism of a valuation ring $\mathfrak{O}$ of L, we see that the image of $1/x$ is 0 and hence that x cannot be in this valuation ring.

Conversely, assume that x is integral over $\mathfrak{o}$, and let

$$x^n + a_{n-1}x^{n-1} + \cdots + a_0 = 0$$

be an integral equation for x with coefficients in $\mathfrak{o}$. Let $\mathfrak{O}$ be any valuation ring of L lying above $\mathfrak{o}$. Suppose $x \notin \mathfrak{O}$. Let φ be the place given by the canonical homomorphism of $\mathfrak{O}$ modulo its maximal ideal. Then $\varphi(x) = \infty$ so $\varphi(1/x) = 0$. Divide the above equation by x^n, and apply φ. Then each term except the first maps to 0 under φ, so we get $\varphi(1) = 0$, a contradiction which proves the proposition.

Proposition 3.6. *Let A be a ring contained in a field L. An element x of L is integral over A if and only if x lies in every valuation ring $\mathfrak{O}$ of L containing A. In terms of places, x is integral over A if and only if every place of L finite on A is finite on x.*

Proof. Assume that every place finite on A is finite on x. We may assume $x \neq 0$. If $1/x$ is a unit in $A[1/x]$ then we can write

$$x = c_0 + c_1(1/x) + \cdots + c_{n-1}(1/x)^{n-1}$$

with $c_i \in A$ and some n. Multiplying by x^{n-1} we conclude that x is integral over A. If $1/x$ is not a unit in $A[1/x]$, then $1/x$ generates a proper principal ideal. By Zorn's lemma this ideal is contained in a maximal ideal $\mathfrak{M}$. The homomorphism $A[1/x] \to A[1/x]/\mathfrak{M}$ can be extended to a place which is a finite on A but maps

$1/x$ on 0, so x on ∞, which contradicts the possibility that $1/x$ is not a unit in $A[1/x]$ and proves that x is integral over A. The converse implication is proved just as in the second part of Proposition 3.5.

Remark. Let K be a subfield of L and let $x \in L$. Then x is integral over K if and only if x is algebraic over K. So if a place φ of L is finite on K, and x is algebraic over K, then φ is finite on $K(x)$. Of course this is a trivial case of the integrality criterion which can be seen directly. Let

$$x^n + a_{n-1}x^{n-1} + \cdots + a_0 = 0$$

be the irreducible equation for x over K. Suppose $x \neq 0$. Then $a_0 \neq 0$. Hence $\varphi(x) \neq 0$ immediately from the equation, so φ is an isomorphism of $K(x)$ on its image.

The next result is a generalization whose technique of proof can also be used in Exercise 1 of Chapter IX (the Hilbert-Zariski theorem).

Theorem 3.7. General Integrality Criterion. *Let A be an entire ring. Let $z_1, \ldots, z_m$ be elements of some extension field of its quotient field K. Assume that each z_s ($s = 1, \ldots, m$) satisfies a polynomial relation*

$$z_s^{d_s} + g_s(z_1, \ldots, z_m) = 0$$

where $g_s(Z_1, \ldots, Z_m) \in A[Z_1, \ldots, Z_m]$ is a polynomial of total degree $< d_s$, and that any pure power of Z_s occuring with non-zero coefficient in g_s occurs with a power strictly less than d_s. Then $z_1, \ldots, z_m$ are integral over A.

Proof. We apply Proposition 3.6. Suppose some z_s is not integral over A. There exists a place φ of K, finite on A, such that $\varphi(z_s) = \infty$ for some s. By Proposition 3.4 we can pick an index s such that $\varphi(z_j/z_s) \neq \infty$ for all j. We divide the polynomial relation of the hypothesis in the lemma by $z_s^{d_s}$ and apply the place. By the hypothesis on g_s, it follows that $\varphi(g_s(z)/z_s^{d_s}) = 0$, whence we get $1 = 0$, a contradiction which proves the theorem.

EXERCISES

1. Let K be a Galois extension of the rationals $\mathbf{Q}$, with group G. Let B be the integral closure of $\mathbf{Z}$ in K, and let $\alpha \in B$ be such that $K = \mathbf{Q}(\alpha)$. Let $f(X) = \mathrm{Irr}(\alpha, \mathbf{Q}, X)$. Let p be a prime number, and assume that f remains irreducible mod p over $\mathbf{Z}/p\mathbf{Z}$. What can you say about the Galois group G? (Artin asked this question to Tate on his qualifying exam.)
2. Let A be an entire ring and K its quotient field. Let t be transcendental over K. If A is integrally closed, show that $A[t]$ is integrally closed.

For the following exercises, you can use §1 of Chapter X.

3. Let A be an entire ring, integrally closed in its quotient field K. Let L be a finite separable extension of K, and let B be the integral closure of A in L. If A is Noetherian, show that B is a finite A-module. [*Hint*: Let $\{\omega_1, \ldots, \omega_n\}$ be a basis of L over K. Multiplying all elements of this basis by a suitable element of A, we may assume without loss of generality that all ω_i are integral over A. Let $\{\omega'_1, \ldots, \omega'_n\}$ be the dual basis relative to the trace, so that $\text{Tr}(\omega_i \omega'_j) = \delta_{ij}$. Write an element α of L integral over A in the form

$$\alpha = b_1\omega'_1 + \cdots + b_n\omega'_n$$

with $b_j \in K$. Taking the trace $\text{Tr}(\alpha\omega_i)$, for $i = 1, \ldots, n$, conclude that B is contained in the finite module $A\omega'_1 + \cdots + A\omega'_n$.] Hence B is Noetherian.

4. The preceding exercise applies to the case when $A = \mathbf{Z}$ and $k = \mathbf{Q}$. Let L be a finite extension of $\mathbf{Q}$ and let $\mathfrak{o}_L$ be the ring of algebraic integers in L. Let $\sigma_1, \ldots, \sigma_n$ be the distinct embeddings of L into the complex numbers. Embedded $\mathfrak{o}_L$ into a Euclidean space by the map

$$\alpha \mapsto (\sigma_1\alpha, \ldots, \sigma_n\alpha).$$

Show that in any bounded region of space, there is only a finite number of elements of $\mathfrak{o}_L$. [*Hint*: The coefficients in an integral equation for α are elementary symmetric functions of the conjugates of α and thus are bounded integers.] Use Exercise 5 of Chapter III to conclude that $\mathfrak{o}_L$ is a free $\mathbf{Z}$-module of dimension $\leqq n$. In fact, show that the dimension is n, a basis of $\mathfrak{o}_L$ over $\mathbf{Z}$ also being a basis of L over $\mathbf{Q}$.

5. Let E be a finite extension of $\mathbf{Q}$, and let $\mathfrak{o}_E$ be the ring of algebraic integers of E. Let U be the group of units of $\mathfrak{o}_E$. Let $\sigma_1, \ldots, \sigma_n$ be the distinct embeddings of E into $\mathbf{C}$. Map U into a Euclidean space, by the map

$$l: \alpha \mapsto (\log|\sigma_1\alpha|, \ldots, \log|\sigma_n\alpha|).$$

Show that $l(U)$ is a free abelian group, finitely generated, by showing that in any finite region of space, there is only a finite number of elements of $l(U)$. Show that the kernel of l is a finite group, and is therefore the group of roots of unity in E. Thus U itself is a finitely generated abelian group.

6. Generalize the results of §2 to infinite Galois extensions, especially Propositions 2.1 and 2.5, using Zorn's lemma.

7. **Dedekind rings.** Let $\mathfrak{o}$ be an entire ring which is Noetherian, integrally closed, and such that every non-zero prime ideal is maximal. Define a fractional ideal $\mathfrak{a}$ to be an $\mathfrak{o}$-submodule $\neq 0$ of the quotient field K such that there exists $c \in \mathfrak{o}$, $c \neq 0$ for which $c\mathfrak{a} \subset \mathfrak{o}$. Prove that the fractional ideals form a group under multiplication. Hint following van der Waerden: Prove the following statements in order:
   (a) Given an ideal $\mathfrak{a} \neq 0$ in $\mathfrak{o}$, there exists a product of prime ideals $\mathfrak{p}_1 \cdots \mathfrak{p}_r \subset \mathfrak{a}$.
   (b) Every maximal ideal $\mathfrak{p}$ is invertible, i.e. if we let $\mathfrak{p}^{-1}$ be the set of elements $x \in K$ such that $x\mathfrak{p} \subset \mathfrak{o}$, then $\mathfrak{p}^{-1}\mathfrak{p} = \mathfrak{o}$.
   (c) Every non-zero ideal is invertible, by a fractional ideal. (Use the Noetherian property that if this is not true, there exists a maximal non-invertible ideal $\mathfrak{a}$, and get a contradiction.)

8. Using prime ideals instead of prime numbers for a Dedekind ring A, define the notion of content as in the Gauss lemma, and prove that if $f(X)$, $g(X) \in A[X]$ are polynomials of degree $\geqq 0$ with coefficients in A, then $\text{cont}(fg) = \text{cont}(f)\text{cont}(g)$. Also if K is the quotient field of A, prove the same statement for $f, g \in K[X]$.

9. Let A be an entire ring, integrally closed. Let B be entire, integral over A. Let Q_1, Q_2 be prime ideals of B with $Q_1 \supset Q_2$ but $Q_1 \neq Q_2$. Let $P_i = Q_i \cap A$. Show that $P_1 \neq P_2$.

10. Let n be a positive integer and let ζ, ζ' be primitive n-th roots of unity.
    (a) Show that $(1 - \zeta)/(1 - \zeta')$ is an algebraic integer.
    (b) If $n \geqq 6$ is divisible by at least two primes, show that $1 - \zeta$ is a unit in the ring $\mathbf{Z}[\zeta]$.

11. Let p be a prime and ζ a primitive p-th root of unity. Show that there is a principal ideal J in $\mathbf{Z}[\zeta]$ such that $J^{p-1} = (p)$ (the principal ideal generated by p).

Symmetric Polynomials

12. Let F be a field of characteristic 0. Let $t_1, \ldots, t_n$ be algebraically independent over F. Let $s_1, \ldots, s_n$ be the elementary symmetric functions. Then $R = F[t_1, \ldots, t_n]$ is an integral extension of $S = F[s_1, \ldots, s_n]$, and actually is its integral closure in the rational field $F(t_1, \ldots, t_n)$. Let W be the group of permutation of the variables $t_1, \ldots, t_n$.
    (a) Show that $S = R^W$ is the fixed subring of R under W.
    (b) Show that the elements $t_1^{r_1} \cdots t_n^{r_n}$ with $0 \leqq r_i \leqq n - i$ form a basis of R over S, so in particular, R is free over S.

I am told that the above basis is due to Kronecker. There is a much more interesting basis, which can be defined as follows.

Let $\partial_1, \ldots, \partial_n$ be the partial derivatives with respect to $t_1, \ldots, t_n$, so $\partial_i = \partial/\partial t_i$. Let $P \in F[t] = F[t_1, \ldots, t_n]$. Substituting ∂_i for t_i $(i = 1, \ldots, n)$ gives a partial differential operator $P(\partial) = P(\partial_1, \ldots, \partial_n)$ on R. An element of S can also be viewed as an element of R. Let $Q \in R$. We say that Q is **W-harmonic** if $P(\partial)Q = 0$ for all symmetric polynomials $P \in S$ with 0 constant term. It can be shown that the W-harmonic polynomials form a finite dimensional space. Furthermore, if $\{H_1, \ldots, H_N\}$ is a basis for this space over F, then it is also a basis for R over S. This is a special case of a general theorem of Chevalley. See [La 99b], where the special case is worked out in detail.

CHAPTER VIII

Transcendental Extensions

Both for their own sake and for applications to the case of finite extensions of the rational numbers, one is led to deal with ground fields which are function fields, i.e. finitely generated over some field k, possibly by elements which are not algebraic. This chapter gives some basic properties of such fields.

§1. TRANSCENDENCE BASES

Let K be an extension field of a field k. Let S be a subset of K. We recall that S (or the elements of S) is said to be algebraically independent over k, if whenever we have a relation

$$0 = \sum a_{(\nu)} M_{(\nu)}(S) = \sum a_{(\nu)} \prod_{x \in S} x^{\nu(x)}$$

with coefficients $a_{(\nu)} \in k$, almost all $a_{(\nu)} = 0$, then we must necessarily have all $a_{(\nu)} = 0$.

We can introduce an ordering among algebraically independent subsets of K, by ascending inclusion. These subsets are obviously inductively ordered, and thus there exist maximal elements. If S is a subset of K which is algebraically independent over k, and if the cardinality of S is greatest among all such subsets, then we call this cardinality the **transcendence degree or dimension** of K over k. Actually, we shall need to distinguish only between finite transcendence degree or infinite transcendence degree. We observe that

the notion of transcendence degree bears to the notion of algebraic independence the same relation as the notion of dimension bears to the notion of linear independence.

We frequently deal with families of elements of K, say a family $\{x_i\}_{i \in I}$, and say that such a family is algebraically independent over k if its elements are distinct (in other words, $x_i \neq x_j$ if $i \neq j$) and if the set consisting of the elements in this family is algebraically independent over k.

A subset S of K which is algebraically independent over k and is maximal with respect to the inclusion ordering will be called a **transcendence base** of K over k. From the maximality, it is clear that if S is a transcendence base of K over k, then K is algebraic over $k(S)$.

Theorem 1.1. *Let K be an extension of a field k. Any two transcendence bases of K over k have the same cardinality. If Γ is a subset of K such that K is algebraic over $k(\Gamma)$, and S is a subset of Γ which is algebraically independent over k, then there exists a transcendence base of K over k such that $S \subset \mathcal{B} \subset \Gamma$.*

Proof. We shall prove that if there exists one finite transcendence base, say $\{x_1, \ldots, x_m\}$, $m \geqq 1$, m minimal, then any other transcendence base must also have m elements. For this it will suffice to prove: If $w_1, \ldots, w_n$ are elements of K which are algebraically independent over k then $n \leqq m$ (for we can then use symmetry). By assumption, there exists a non-zero irreducible polynomial f_1 in $m + 1$ variables with coefficients in k such that

$$f_1(w_1, x_1, \ldots, x_m) = 0.$$

After renumbering $x_1, \ldots, x_m$ we may write $f_1 = \sum g_j(w_1, x_2, \ldots, x_m) x_1^j$ with some $g_N \neq 0$ with some $N \geqq 1$. No irreducible factor of g_N vanishes on $(w_1, x_2, \ldots, x_n)$, otherwise w_1 would be a root of two distinct irreducible polynomials over $k(x_1, \ldots, x_m)$. Hence x_1 is algebraic over $k(w_1, x_2, \ldots, x_m)$ and $w_1, x_2, \ldots, x_m$ are algebraically independent over k, otherwise the minimality of m would be contradicted. Suppose inductively that after a suitable renumbering of $x_2, \ldots, x_m$ we have found $w_1, \ldots, w_r$ ($r < n$) such that K is algebraic over $k(w_1, \ldots, w_r, x_{r+1}, \ldots, x_m)$. Then there exists a non-zero polynomial f in $m + 1$ variables with coefficients in k such that

$$f(w_{r+1}, w_1, \ldots, w_r, x_{r+1}, \ldots, x_m) = 0.$$

Since the w's are algebraically independent over k, it follows by the same argument as in the first step that some x_j, say x_{r+1}, is algebraic over $k(w_1, \ldots, w_{r+1}, x_{r+2}, \ldots, x_m)$. Since a tower of algebraic extensions is algebraic, it follows that K is algebraic over $k(w_1, \ldots, w_{r+1}, x_{r+2}, \ldots, x_m)$. We can repeat the procedure, and if $n \geqq m$ we can replace all the x's by w's, to see that K is algebraic over $k(w_1, \ldots, w_m)$. This shows that $n \geqq m$ implies $n = m$, as desired.

We have now proved: Either the transcendence degree is finite, and is equal to the cardinality of any transcendence base, or it is infinite, and every transcendence base is infinite. The cardinality statement in the infinite case will be left as an exercise. We shall also leave as an exercise the statement that a set of algebraically independent elements can be completed to a transcendence base, selected from a given set Γ such that K is algebraic over $k(\Gamma)$. (The reader will note the complete analogy of our statements with those concerning linear bases.)

Note. *The preceding section is the only one used in the next chapter. The remaining sections are more technical, especially §3 and §4 which will not be used in the rest of the book. Even §2 and §5 will only be mentioned a couple of times, and so the reader may omit them until they are referred to again.*

§2. NOETHER NORMALIZATION THEOREM

Theorem 2.1. *Let $k[x_1, \ldots, x_n] = k[x]$ be a finitely generated entire ring over a field k, and assume that $k(x)$ has transcendence degree r. Then there exist elements $y_1, \ldots, y_r$ in $k[x]$ such that $k[x]$ is integral over*

$$k[y] = k[y_1, \ldots, y_r].$$

Proof. If $(x_1, \ldots, x_n)$ are already algebraically independent over k, we are done. If not, there is a non-trivial relation

$$\sum a_{(j)} x_1^{j_1} \cdots x_n^{j_n} = 0$$

with each coefficient $a_{(j)} \in k$ and $a_{(j)} \neq 0$. The sum is taken over a finite number of distinct n-tuples of integers $(j_1, \ldots, j_n)$, $j_\nu \geqq 0$. Let $m_2, \ldots, m_n$ be positive integers, and put

$$y_2 = x_2 - x_1^{m_2}, \ldots, y_n = x_n - x_1^{m_n}.$$

Substitute $x_i = y_i + x_1^{m_i}$ $(i = 2, \ldots, n)$ in the above equation. Using vector notation, we put $(m) = (1, m_2, \ldots, m_n)$ and use the dot product $(j)\cdot(m)$ to denote $j_1 + m_2 j_2 + \cdots + m_n j_n$. If we expand the relation after making the above substitution, we get

$$\sum c_{(j)} x_1^{(j)\cdot(m)} + f(x_1, y_2, \ldots, y_n) = 0$$

where f is a polynomial in which no pure power of x_1 appears. We now select d to be a large integer [say greater than any component of a vector (j) such that $c_{(j)} \neq 0$] and take

$$(m) = (1, d, d^2, \ldots, d^n).$$

Then all $(j)\cdot(m)$ are distinct for those (j) such that $c_{(j)} \neq 0$. In this way we obtain an integral equation for x_1 over $k[y_2, \ldots, y_n]$. Since each x_i $(i > 1)$ is integral over $k[x_1, y_2, \ldots, y_n]$, it follows that $k[x]$ is integral over $k[y_2, \ldots, y_n]$. We can now proceed inductively, using the transitivity of integral extensions to shrink the number of y's until we reach an algebraically independent set of y's.

The advantage of the proof of Theorem 2.1 is that it is applicable when k is a finite field. The disadvantage is that it is not linear in $x_1, \ldots, x_n$. We now deal with another technique which leads into certain aspects of algebraic geometry on which we shall comment after the next theorem.

We start again with $k[x_1, \ldots, x_n]$ finitely generated over k and entire. Let (u_{ij}) $(i, j = 1, \ldots, n)$ be algebraically independent elements over $k(x)$, and let $k_u = k(u) = k(u_{ij})_{\text{all } i,j}$. Put

$$y_i = \sum_{j=1}^{n} u_{ij} x_j.$$

This amounts to a generic linear change of coordinates in n-space, to use geometric terminology. Again we let r be the transcendence degree of $k(x)$ over k.

Theorem 2.2. *With the above notation, $k_u[x]$ is integral over $k_u[y_1, \ldots, y_r]$.*

Proof. Suppose some x_i is not integral over $k_u[y_1, \ldots, y_r]$. Then there exists a place φ of $k_u(y)$ finite on $k_u[y_1, \ldots, y_r]$ but taking the value ∞ on some x_i. Using Proposition 3.4 of Chapter VII, and renumbering the indices if necessary, say $\varphi(x_j/x_n)$ is finite for all i. Let $z'_j = \varphi(x_j/x_n)$ for $j = 1, \ldots, n$. Then dividing the equations $y_i = \sum u_{ij} x_j$ by x_n (for $i = 1, \ldots, r$) and applying the place, we get

$$0 = u_{11} z'_1 + u_{12} z'_2 + \cdots + u_{1n},$$
$$\vdots$$
$$0 = u_{r1} z'_1 + u_{r2} z'_2 + \cdots + u_{rn}.$$

The transcendence degree of $k(z')$ over k cannot be r, for otherwise, the place φ would be an isomorphism of $k(x)$ on its image. [Indeed, if, say, $z'_1, \ldots, z'_r$ are algebraically independent and $z_i = x_i/x_n$, then $z_1, \ldots, z_r$ are also algebraically independent, and so form a transcendence base for $k(x)$ over k. Then the place is an isomorphism from $k(z_1, \ldots, z_r)$ to $k(z'_1, \ldots, z'_r)$, and hence is an isomorphism from $k(x)$ to its image.] We then conclude that

$$u_{1n}, \ldots, u_{rn} \in k(u_{ij}, z') \qquad \text{with} \quad i = 1, \ldots, r; \quad j = 1, \ldots, n-1.$$

Hence the transcendence degree of $k(u)$ over k would be $\leqq rn - 1$, which is a contradiction, proving the theorem.

Corollary 2.3. *Let k be a field, and let $k(x)$ be a finitely generated extension of transcendence degree r. There exists a polynomial $P(u) = P(u_{ij}) \in k[u]$ such that if $(c) = (c_{ij})$ is a family of elements $c_{ij} \in k$ satisfying $P(c) \neq 0$, and we let $y'_i = \sum c_{ij}x_j$, then $k[x]$ is integral over $k[y'_1, \ldots, y'_r]$.*

Proof. By Theorem 2.2, each x_i is integral over $k_u[y_1, \ldots, y_r]$. The coefficients of an integral equation are rational functions in k_u. We let $P(u)$ be a common denominator for these rational functions. If $P(c) \neq 0$, then there is a homomorphism

$$\varphi: k(x)[u, P(u)^{-1}] \to k(x)$$

such that $\varphi(u) = (c)$, and such that φ is the identity on $k(x)$. We can apply φ to an integral equation for x_i over $k_u[y]$ to get an integral equation for x_i over $k[y']$, thus concluding the proof.

Remark. After Corollary 2.3, there remains the problem of finding explicitly integral equations for $x_1, \ldots, x_n$ (or $y_{r+1}, \ldots, y_n$) over $k_u[y_1, \ldots, y_r]$. This is an elimination problem, and I have decided to refrain from further involvement in algebraic geometry at this point. But it may be useful to describe the geometric language used to interpret Theorem 2.2 and further results in that line. After the generic change of coordinates, the map

$$(y_1, \ldots, y_n) \mapsto (y_1, \ldots, y_r)$$

is the generic projection of the variety whose coordinate ring is $k[x]$ on affine r-space. This projection is finite, and in particular, the inverse image of a point on affine r-space is finite. Furthermore, if $k(x)$ is separable over k (a notion which will be defined in §4), then the extension $k_u(y)$ is finite separable over $k_u(y_1, \ldots, y_r)$ (in the sense of Chapter V). To determine the degree of this finite extension is essentially Bezout's theorem. Cf. [La 58], Chapter VIII, §6.

The above techniques were created by van der Waerden and Zariski, cf., for instance, also Exercises 5 and 6. These techniques have unfortunately not been completely absorbed in some more recent expositions of algebraic geometry. To give a concrete example: When Hartshorne considers the intersection of a variety and a sufficiently general hyperplane, he does not discuss the "generic" hyperplane (that is, with algebraically independent coefficients over a given ground field), and he assumes that the variety is non-singular from the start (see his Theorem 8.18 of Chapter 8, [Ha 77]). But the description of the intersection can be done without simplicity assumptions, as in Theorem 7 of [La 58], Chapter VII, §6, and the corresponding lemma. Something was lost in discarding the technique of the algebraically independent (u_{ij}).

After two decades when the methods illustrated in Chapter X have been prevalent, there is a return to the more explicit methods of generic constructions using the algebraically independent (u_{ij}) and similar ones for some

applications because part of algebraic geometry and number theory are returning to some problems asking for explicit or effective constructions, with bounds on the degrees of solutions of algebraic equations. See, for instance, [Ph 91–95], [So 90], and the bibliography at the end of Chapter X, §6. Returning to some techniques, however, does not mean abandoning others; it means only expanding available tools.

Bibliography

[Ha 77] R. HARTSHORNE, *Algebraic Geometry*, Springer-Verlag, New York, 1977

[La 58] S. LANG, *Introduction to Algebraic Geometry*, Wiley–Interscience, New York, 1958

[Ph 91–95] P. PHILIPPON, Sur des hauteurs alternatives, I *Math. Ann.* 289 (1991) pp. 255–283; II *Ann. Inst. Fourier* 44 (1994) pp. 1043–1065; III *J. Math. Pures Appl.* 74 (1995) pp. 345–365

[So 90] C. SOULÉ, Géometrie d'Arakelov et théorie des nombres transcendants, Asterisque 198–200 (1991) pp. 355–371

§3. LINEARLY DISJOINT EXTENSIONS

In this section we discuss the way in which two extensions K and L of a field k behave with respect to each other. We assume that all the fields involved are contained in one field Ω, assumed algebraically closed.

K is said to be **linearly disjoint from L over k** if every finite set of elements of K that is linearly independent over k is still such over L.

The definition is unsymmetric, but we prove right away that the property of being linearly disjoint is actually symmetric for K and L. Assume K linearly disjoint from L over k. Let $y_1, \dots, y_n$ be elements of L linearly independent over k. Suppose there is a non-trivial relation of linear dependence over K,

$$x_1 y_1 + x_2 y_2 + \cdots + x_n y_n = 0. \tag{1}$$

Say $x_1, \dots, x_r$ are linearly independent over k, and $x_{r+1}, \dots, x_n$ are linear combinations $x_i = \sum_{\mu=1}^{r} a_{i\mu} x_\mu$, $i = r+1, \dots, n$. We can write the relation (1) as follows:

$$\sum_{\mu=1}^{r} x_\mu y_\mu + \sum_{i=r+1}^{n} \left(\sum_{\mu=1}^{r} a_{i\mu} x_\mu \right) y_i = 0$$

and collecting terms, after inverting the second sum, we get

$$\sum_{\mu=1}^{r} \left(y_\mu + \sum_{i=r+1}^{n} (a_{i\mu} y_i) \right) x_\mu = 0.$$

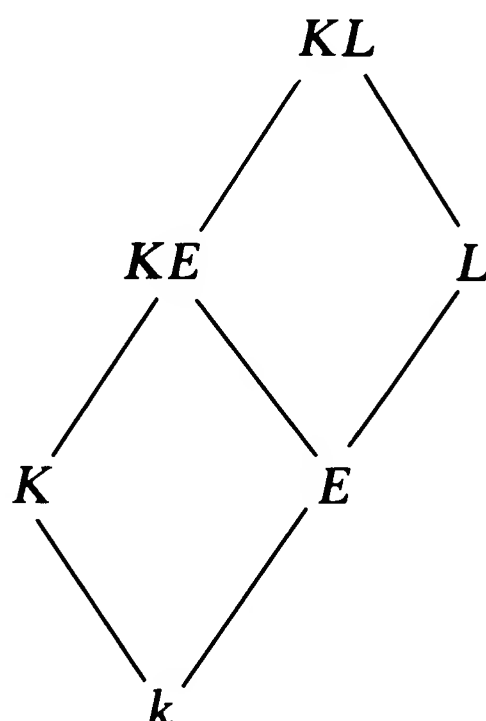
The y's are linearly independent over k, so the coefficients of x_μ are $\neq 0$. This contradicts the linear disjointness of K and L over k.

We now give two criteria for linear disjointness.

Criterion 1. Suppose that K is the quotient field of a ring R and L the quotient field of a ring S. To test whether L and K are linearly disjoint, it suffices to show that if elements $y_1, \dots, y_n$ of S are linearly independent over k, then there is no linear relation among the y's with coefficients in R. Indeed, if elements $y_1, \dots, y_n$ of L are linearly independent over k, and if there is a relation $x_1 y_1 + \cdots + x_n y_n = 0$ with $x_i \in K$, then we can select y in S and x in R such that $xy \neq 0$, $yy_i \in S$ for all i, and $xx_i \in R$ for all i. Multiplying the relation by xy gives a linear dependence between elements of R and S. However, the yy_i are obviously linearly independent over k, and this proves our criterion.

Criterion 2. Again let R be a subring of K such that K is its quotient field and R is a vector space over k. Let $\{u_\alpha\}$ be a basis of R considered as a vector space over k. To prove K and L linearly disjoint over k, it suffices to show that the elements $\{u_\alpha\}$ of this basis remain linearly independent over L. Indeed, suppose this is the case. Let $x_1, \dots, x_m$ be elements of R linearly independent over k. They lie in a finite dimension vector space generated by some of the u_α, say $u_1, \dots, u_n$. They can be completed to a basis for this space over k. Lifting this vector space of dimension n over L, it must conserve its dimension because the u's remain linearly independent by hypothesis, and hence the x's must also remain linearly independent.

Proposition 3.1. *Let K be a field containing another field k, and let $L \supset E$ be two other extensions of k. Then K and L are linearly disjoint over k if and only if K and E are linearly disjoint over k and KE, L are linearly disjoint over E.*



Proof. Assume first that K, E are linearly disjoint over k, and KE, L are linearly disjoint over E. Let $\{\kappa\}$ be a basis of K as vector space over k (we use the elements of this basis as their own indexing set), and let $\{\alpha\}$ be a basis of E over k. Let $\{\lambda\}$ be a basis of L over E. Then $\{\alpha\lambda\}$ is a basis of L over k. If K and L are not linearly disjoint over k, then there exists a relation

$$\sum_{\lambda,\alpha}\left(\sum_{\kappa} c_{\kappa\lambda\alpha}\kappa\right)\lambda\alpha = 0 \qquad \text{with some } c_{\kappa\lambda\alpha} \neq 0,\ c_{\kappa\lambda\alpha} \in k.$$

Changing the order of summation gives

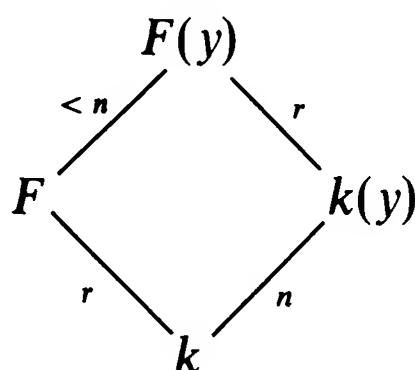
$$\sum_{\lambda}\left(\sum_{\kappa,\lambda} c_{\kappa\lambda\alpha}\kappa\alpha\right)\lambda = 0$$

contradicting the linear disjointness of L and KE over E.

Conversely, assume that K and L are linearly disjoint over k. Then *a fortiori*, K and E are also linearly disjoint over k, and the field KE is the quotient field of the ring $E[K]$ generated over E by all elements of K. This ring is a vector space over E, and a basis for K over k is also a basis for this ring $E[K]$ over E. With this remark, and the criteria for linear disjointness, we see that it suffices to prove that the elements of such a basis remain linearly independent over L. At this point we see that the arguments given in the first part of the proof are reversible. We leave the formalism to the reader.

We introduce another notion concerning two extensions K and L of a field k. We shall say that K is **free from L over k** if every finite set of elements of K algebraically independent over k remains such over L. If (x) and (y) are two sets of elements in Ω, we say that they are **free over k** (or **independent over k**) if $k(x)$ and $k(y)$ are free over k.

Just as with linear disjointness, our definition is unsymmetric, and we prove that the relationship expressed therein is actually symmetric. Assume therefore that K is free from L over k. Let $y_1, \ldots, y_n$ be elements of L, algebraically independent over k. Suppose they become dependent over K. They become so in a subfield F of K finitely generated over k, say of transcendence degree r over k. Computing the transcendence degree of $F(y)$ over k in two ways gives a contradiction (cf. Exercise 5).



Proposition 3.2. *If K and L are linearly disjoint over k, then they are free over k.*

Proof. Let $x_1, \dots, x_n$ be elements of K algebraically independent over k. Suppose they become algebraically dependent over L. We get a relation

$$\sum y_a M_\alpha(x) = 0$$

between monomials $M_\alpha(x)$ with coefficients y_α in L. This gives a linear relation among the $M_\alpha(x)$. But these are linearly independent over k because the x's are assumed algebraically independent over k. This is a contradiction.

Proposition 3.3. *Let L be an extension of k, and let $(u) = (u_1, \dots, u_r)$ be a set of quantities algebraically independent over L. Then the field $k(u)$ is linearly disjoint from L over k.*

Proof. According to the criteria for linear disjointness, it suffices to prove that the elements of a basis for the ring $k[u]$ that are linearly independent over k remain so over L. In fact the monomials $M(u)$ give a basis of $k[u]$ over k. They must remain linearly independent over L, because as we have seen, a linear relation gives an algebraic relation. This proves our proposition.

Note finally that the property that two extensions K and L of a field k are linearly disjoint or free is of finite type. To prove that they have either property, it suffices to do it for all subfields K_0 and L_0 of K and L respectively which are finitely generated over k. This comes from the fact that the definitions involve only a finite number of quantities at a time.

§4. SEPARABLE AND REGULAR EXTENSIONS

Let K be a finitely generated extension of k, $K = k(x)$. We shall say that it is **separably generated** if we can find a transcendence basis $(t_1, \dots, t_r)$ of K/k such that K is separably algebraic over $k(t)$. Such a transcendence base is said to be a **separating transcendence base** for K over k.

We always denote by p the characteristic if it is not 0. The field obtained from k by adjoining all p^m-th roots of all elements of k will be denoted by k^{1/p^m}. The compositum of all such fields for $m = 1, 2, \dots$, is denoted by k^{1/p^∞}.

Proposition 4.1. *The following conditions concerning an extension field K of k are equivalent:*

(i) *K is linearly disjoint from k^{1/p^∞}.*
(ii) *K is linearly disjoint from k^{1/p^m} for some m.*

(iii) *Every subfield of K containing k and finitely generated over k is separably generated.*

Proof. It is obvious that (i) implies (ii). In order to prove that (ii) implies (iii), we may clearly assume that K is finitely generated over k, say

$$K = k(x) = k(x_1, \dots, x_n).$$

Let the transcendence degree of this extension be r. If $r = n$, the proof is complete. Otherwise, say $x_1, \dots, x_r$ is a transcendence base. Then x_{r+1} is algebraic over $k(x_1, \dots, x_r)$. Let $f(X_1, \dots, X_{r+1})$ be a polynomial of lowest degree such that

$$f(x_1, \dots, x_{r+1}) = 0.$$

Then f is irreducible. We contend that not all x_i $(i = 1, \dots, r + 1)$ appear to the p-th power throughout. If they did, we could write $f(X) = \sum c_\alpha M_\alpha(X)^p$ where $M_\alpha(X)$ are monomials in $X_1, \dots, X_{r+1}$ and $c_\alpha \in k$. This would imply that the $M_\alpha(x)$ are linearly dependent over $k^{1/p}$ (taking the p-th root of the equation $\sum c_\alpha M_\alpha(x)^p = 0$). However, the $M_\alpha(x)$ are linearly independent over k (otherwise we would get an equation for $x_1, \dots, x_{r+1}$ of lower degree) and we thus get a contradiction to the linear disjointness of $k(x)$ and $k^{1/p}$. Say X_1 does not appear to the p-th power throughout, but actually appears in $f(X)$. We know that $f(X)$ is irreducible in $k[X_1, \dots, X_{r+1}]$ and hence $f(x) = 0$ is an irreducible equation for x_1 over $k(x_2, \dots, x_{r+1})$. Since X_1 does not appear to the p-th power throughout, this equation is a separable equation for x_1 over $k(x_2, \dots., x_{r+1})$, in other words, x_1 is separable algebraic over $k(x_2, \dots, x_{r+1})$. From this it follows that it is separable algebraic over $k(x_2, \dots, x_n)$. If $(x_2, \dots, x_n)$ is a transcendence base, the proof is complete. If not, say that x_2 is separable over $k(x_3, \dots, x_n)$. Then $k(x)$ is separable over $k(x_3, \dots, x_n)$. Proceeding inductively, we see that the procedure can be continued until we get down to a transcendence base. This proves that (ii) implies (iii). It also proves that a separating transcendence base for $k(x)$ over k can be selected from the given set of generators (x).

To prove that (iii) implies (i) we may assume that K is finitely generated over k. Let (u) be a transcendence base for K over k. Then K is separably algebraic over $k(u)$. By Proposition 3.3, $k(u)$ and k^{1/p^∞} are linearly disjoint. Let $L = k^{1/p^\infty}$. Then $k(u)L$ is purely inseparable over $k(u)$, and hence is linearly disjoint from K over $k(u)$ by the elementary theory of finite algebraic extensions. Using Proposition 3.1, we conclude that K is linearly disjoint from L over k, thereby proving our theorem.

An extension K of k satisfying the conditions of Proposition 4.1 is called **separable**. This definition is compatible with the use of the word for algebraic extensions.

The first condition of our theorem is known as **MacLane's criterion**. It has the following immediate corollaries.

Corollary 4.2. *If K is separable over k, and E is a subfield of K containing k, then E is separable over k.*

Corollary 4.3. *Let E be a separable extension of k, and K a separable extension of E. Then K is a separable extension of k.*

Proof. Apply Proposition 3.1 and the definition of separability.

Corollary 4.4. *If k is perfect, every extension of k is separable.*

Corollary 4.5. *Let K be a separable extension of k, and free from an extension L of k. Then KL is a separable extension of L.*

Proof. An element of KL has an expression in terms of a finite number of elements of K and L. Hence any finitely generated subfield of KL containing L is contained in a composite field FL, where F is a subfield of K finitely generated over k. By Corollary 4.2, we may assume that K is finitely generated over k. Let (t) be a transcendence base of K over k, so K is separable algebraic over $k(t)$. By hypothesis, (t) is a transcendence base of KL over L, and since every element of K is separable algebraic over $k(t)$, it is also separable over $L(t)$. Hence KL is separably generated over L. This proves the corollary.

Corollary 4.6. *Let K and L be two separable extensions of k, free from each other over k. Then KL is separable over k.*

Proof. Use Corollaries 4.5 and 4.3.

Corollary 4.7. *Let K, L be two extensions of k, linearly disjoint over k. Then K is separable over k if and only if KL is separable over L.*

Proof. If K is not separable over k, it is not linearly disjoint from $k^{1/p}$ over k, and hence *a fortiori* it is not linearly disjoint from $Lk^{1/p}$ over k. By Proposition 4.1, this implies that KL is not linearly disjoint from $Lk^{1/p}$ over L, and hence that KL is not separable over L. The converse is a special case of Corollary 4.5, taking into account that linearly disjoint fields are free.

We conclude our discussion of separability with two results. The first one has already been proved in the first part of Proposition 4.1, but we state it here explicitly.

Proposition 4.8. *If K is a separable extension of k, and is finitely generated, then a separating transcendence base can be selected from a given set of generators.*

To state the second result we denote by K^{p^m} the field obtained from K by raising all elements of K to the p^m-th power.

Proposition 4.9. *Let K be a finitely generated extension of a field k. If $K^{p^m}k = K$ for some m, then K is separably algebraic over k. Conversely, if K is separably algebraic over k, then $K^{p^m}k = K$ for all m.*

Proof. If K/k is separably algebraic, then the conclusion follows from the elementary theory of finite algebraic extensions. Conversely, if K/k is finite algebraic but not separable, then the maximal separable extension of k in K cannot be all of K, and hence $K^p k$ cannot be equal to K. Finally, if there exists an element t of K transcendental over k, then $k(t^{1/p^m})$ has degree p^m over $k(t)$, and hence there exists a t such that t^{1/p^m} does not lie in K. This proves our proposition.

There is a class of extensions which behave particularly well from the point of view of changing the ground field, and are especially useful in algebraic geometry. We put some results together to deal with such extensions. Let K be an extension of a field k, with algebraic closure K^a. We claim that the following two conditions are equivalent:

REG 1. k is algebraically closed in K (i.e. every element of K algebraic over k lies in k), and K is separable over k.

REG 2. K is linearly disjoint from k^a over k.

We show the equivalence. Assume **REG 2**. By Proposition 4.1, we know that K is separably generated over k. It is obvious that k must be algebraically closed in K. Hence **REG 2** implies **REG 1**. To prove the converse we need a lemma.

Lemma 4.10. *Let k be algebraically closed in extension K. Let x be some element of an extension of K, but algebraic over k. Then $k(x)$ and K are linearly disjoint over k, and $[k(x) : k] = [K(x) : K]$.*

Proof. Let $f(X)$ be the irreducible polynomial for x over k. Then f remains irreducible over K; otherwise, its factors would have coefficients algebraic over k, hence in k. Powers of x form a basis of $k(x)$ over k, hence the same powers form a basis of $K(x)$ over K. This proves the lemma.

To prove **REG 2** from **REG 1**, we may assume without loss of generality that K is finitely generated over k, and it suffices to prove that K is linearly disjoint from an arbitrary finite algebraic extension L of k. If L is separable algebraic over k, then it can be generated by one primitive element, and we can apply Lemma 4.10.

More generally, let E be the maximal separable subfield of L containing k. By Proposition 3.1, we see that it suffices to prove that KE and L are linearly disjoint over E. Let (t) be a separating transcendence base for K over k. Then K is separably algebraic over $k(t)$. Furthermore, (t) is also a separating transcendence base for KE over E, and KE is separable algebraic

over $E(t)$. Thus KE is separable over E, and by definition KE is linearly disjoint from L over K because L is purely inseparable over E. This proves that **REG 1** implies **REG 2**.

Thus we can define an extension K of k to be **regular** if it satisfies either one of the equivalent conditions **REG 1** or **REG 2**.

Proposition 4.11.

(a) *Let K be a regular extension of k, and let E be a subfield of K containing k. Then E is regular over k.*
(b) *Let E be a regular extension of k, and K a regular extension of E. Then K is a regular extension of k.*
(c) *If k is algebraically closed, then every extension of k is regular.*

Proof. Each assertion is immediate from the definition conditions **REG 1** and **REG 2**.

Theorem 4.12. *Let K be a regular extension of k, let L be an arbitrary extension of k, both contained in some larger field, and assume that K, L are free over k. Then K, L are linearly disjoint over k.*

Proof (Artin). Without loss of generality, we may assume that K is finitely generated over k. Let $x_1, \ldots, x_n$ be elements of K linearly independent over k. Suppose we have a relation of linear dependence

$$x_1 y_1 + \cdots + x_n y_n = 0$$

with $y_i \in L$. Let φ be a k^a-valued place of L over k. Let (t) be a transcendence base of K over k. By hypothesis, the elements of (t) remain algebraically independent over L, and hence φ can be extended to a place of KL which is identity on $k(t)$. This place must then be an isomorphism of K on its image, because K is a finite algebraic extension of $k(t)$ (remark at the end of Chapter VII, §3). After a suitable isomorphism, we may take a place equivalent to φ which is the identity on K. Say $\varphi(y_i/y_n)$ is finite for all i (use Proposition 3.4 of Chapter VII). We divide the relation of linear dependence by y_n and apply φ to get $\sum x_i \varphi(y_i/y_n) = 0$, which gives a linear relation among the x_i with coefficients in k^a, contradicting the linear disjointness. This proves the theorem.

Theorem 4.13. *Let K be a regular extension of k, free from an extension L of k over k. Then KL is a regular extension of L.*

Proof. From the hypothesis, we deduce that K is free from the algebraic closure L^a of L over k. By Theorem 4.12, K is linearly disjoint from L^a over k. By Proposition 3.1, KL is linearly disjoint from L^a over L, and hence KL is regular over L.

Corollary 4.14. *Let K, L be regular extensions of k, free from each other over k. Then KL is a regular extension of k.*

Proof. Use Corollary 4.13 and Proposition 4.11(b).

Theorem 4.13 is one of the main reasons for emphasizing the class of regular extensions: they remain regular under arbitrary base change of the ground field k. Furthermore, Theorem 4.12 in the background is important in the study of polynomial ideals as in the next section, and we add some remarks here on its implications. We now assume that the reader is acquainted with the most basic properties of the tensor product (Chapter XVI, §1 and §2).

Corollary 4.15. *Let $K = k(x)$ be a finitely generated regular extension, free from an extension L of k, and both contained in some larger field. Then the natural k-algebra homomorphism*

$$L \otimes_k k[x] \to L[x]$$

is an isomorphism.

Proof. By Theorem 4.12 the homomorphism is injective, and it is obviously surjective, whence the corollary follows.

Corollary 4.16. *Let $k(x)$ be a finitely generated regular extension, and let $\mathfrak{p}$ be the prime ideal in $k[X]$ vanishing on (x), that is, consisting of all polynomials $f(X) \in k[X]$ such that $f(x) = 0$. Let L be an extension of k, free from $k(x)$ over k. Let $\mathfrak{p}_L$ be the prime ideal in $L[X]$ vanishing on (x). Then $\mathfrak{p}_L = \mathfrak{p}L[X]$, that is $\mathfrak{p}_L$ is the ideal generated by $\mathfrak{p}$ in $L[X]$, and in particular, this ideal is prime.*

Proof. Consider the exact sequence

$$0 \to \mathfrak{p} \to k[X] \to k[x] \to 0.$$

Since we are dealing with vector spaces over a field, the sequence remains exact when tensored with any k-space, so we get an exact sequence

$$0 \to L \otimes_k \mathfrak{p} \to L[X] \to L \otimes_k k[x] \to 0.$$

By Corollary 4.15, we know that $L \otimes_k k[x] \approx L[x]$, and the image of $L \otimes_k \mathfrak{p}$ in $L[X]$ is $\mathfrak{p}L[X]$, so the lemma is proved.

Corollary 4.16 shows another aspect whereby regular extensions behave well under extension of the base field, namely the way the prime ideal $\mathfrak{p}$ remains prime under such extensions.

§5. DERIVATIONS

A **derivation** D of a ring R is a mapping $D\colon R \to R$ of R into itself which is linear and satisfies the ordinary rule for derivatives, i.e.,

$$D(x + y) = Dx + Dy \qquad \text{and} \qquad D(xy) = xDy + yDx.$$

As an example of derivations, consider the polynomial ring $k[X]$ over a field k. For each variable X_i, the partial derivative $\partial/\partial X_i$ taken in the usual manner is a derivation of $k[X]$.

Let R be an entire ring and let K be its quotient field. Let $D\colon R \to R$ be a derivation. Then D extends uniquely to a derivation of K, by defining

$$D(u/v) = \frac{vDu - uDv}{v^2}.$$

It is immediately verified that the expression on the right-hand side is independent of the way we represent an element of K as u/v ($u, v \in R$), and satisfies the conditions defining a derivation.

Note. In this section, we shall discuss derivations of fields. For derivations in the context of rings and modules, see Chapter XIX, §3.

A derivation of a field K is **trivial** if $Dx = 0$ for all $x \in K$. It is trivial **over a subfield** k of K if $Dx = 0$ for all $x \in k$. A derivation is always trivial over the prime field: One sees that

$$D(1) = D(1 \cdot 1) = 2D(1),$$

whence $D(1) = 0$.

We now consider the problem of extending derivations. Let

$$L = K(x) = K(x_1, \ldots, x_n)$$

be a finitely generated extension. If $f \in K[X]$, we denote by $\partial f/\partial x_i$ the polynomials $\partial f/\partial X_i$ evaluated at (x). Given a derivation D on K, does there exist a derivation D^* on L coinciding with D on K? If $f(X) \in K[X]$ is a polynomial vanishing on (x), then any such D^* must satisfy

$$0 = D^*f(x) = f^D(x) + \sum (\partial f/\partial x_i)D^*x_i, \tag{1}$$

where f^D denotes the polynomial obtained by applying D to all coefficients of f. Note that if relation (1) is satisfied for every element in a finite set of generators of the ideal in $K[X]$ vanishing on (x), then (1) is satisfied by every polynomial of this ideal. This is an immediate consequence of the rules for derivations. The preceding ideal will also be called the ideal determined by (x) in $K[X]$.

The above necessary condition for the existence of a D^* turns out to be sufficient.

Theorem 5.1. *Let D be a derivation of a field K. Let*

$$(x) = (x_1, \ldots, x_n)$$

be a finite family of elements in an extension of K. Let $\{f_\alpha(X)\}$ be a set of generators for the ideal determined by (x) in $K[X]$. Then, if (u) is any set of elements of $K(x)$ satisfying the equations

$$0 = f_\alpha^D(x) + \sum (\partial f_\alpha / \partial x_i) u_i,$$

there is one and only one derivation D^* of $K(x)$ coinciding with D on K, and such that $D^* x_i = u_i$ for every i.

Proof. The necessity has been shown above. Conversely, if $g(x)$, $h(x)$ are in $K[x]$, and $h(x) \neq 0$, one verifies immediately that the mapping D^* defined by the formulas

$$D^* g(x) = g^D(x) + \sum \frac{\partial g}{\partial x_i} u_i,$$

$$D^*(g/h) = \frac{h D^* g - g D^* h}{h^2},$$

is well defined and is a derivation of $K(x)$.

Consider the special case where (x) consists of one element x. Let D be a given derivation on K.

Case 1. x is separable algebraic over K. Let $f(X)$ be the irreducible polynomial satisfied by x over K. Then $f'(x) \neq 0$. We have

$$0 = f^D(x) + f'(x)u,$$

whence $u = -f^D(x)/f'(x)$. Hence D extends to $K(x)$ uniquely. If D is trivial on K, then D is trivial on $K(x)$.

Case 2. x is transcendental over K. Then D extends, and u can be selected arbitrarily in $K(x)$.

Case 3. x is purely inseparable over K, so $x^p - a = 0$, with $a \in K$. Then D extends to $K(x)$ if and only if $Da = 0$. In particular if D is trivial on K, then u can be selected arbitrarily.

Proposition 5.2. *A finitely generated extension $K(x)$ over K is separable algebraic if and only if every derivation D of $K(x)$ which is trivial on K is trivial on $K(x)$.*

Proof. If $K(x)$ is separable algebraic over K, this is Case 1. Conversely, if it is not, we can make a tower of extensions between K and $K(x)$, such

that each step is covered by one of the three above cases. At least one step will be covered by Case 2 or 3. Taking the uppermost step of this latter type, one sees immediately how to construct a derivation trivial on the bottom and nontrivial on top of the tower.

Proposition 5.3. *Given K and elements $(x) = (x_1, \ldots, x_n)$ in some extension field, assume that there exist n polynomials $f_i \in K[X]$ such that:*

(i) $f_i(x) = 0$, *and*
(ii) $\det(\partial f_i/\partial x_j) \neq 0$.

Then (x) is separably algebraic over K.

Proof. Let D be a derivation on $K(x)$, trivial on K. Having $f_i(x) = 0$ we must have $Df_i(x) = 0$, whence the Dx_i satisfy n linear equations such that the coefficient matrix has non-zero determinant. Hence $Dx_i = 0$, so D is trivial on $K(x)$. Hence $K(x)$ is separable algebraic over K by Proposition 5.2.

The following proposition will follow directly from Cases 1 and 2.

Proposition 5.4. *Let $K = k(x)$ be a finitely generated extension of k. An element z of K is in K^pk if and only if every derivation D of K over k is such that $Dz = 0$.*

Proof. If z is in K^pk, then it is obvious that every derivation D of K over k vanishes on z. Conversely, if $z \notin K^pk$, then z is purely inseparable over K^pk, and by Case 3 of the extension theorem, we can find a derivation D trivial on K^pk such that $Dz = 1$. This derivation is at first defined on the field $K^pk(z)$. One can extend it to K as follows. Suppose there is an element $w \in K$ such that $w \notin K^pk(z)$. Then $w^p \in K^pk$, and D vanishes on w^p. We can then again apply Case 3 to extend D from $K^pk(z)$ to $K^pk(z, w)$. Proceeding stepwise, we finally reach K, thus proving our proposition.

The derivations D of a field K form a vector space over K if we define zD for $z \in K$ by $(zD)(x) = zDx$.

Let K be a finitely generated extension of k, of dimension r over k. We denote by $\mathfrak{D}$ the K-vector space of derivations D of K over k (derivations of K which are trivial on k). For each $z \in K$, we have a pairing

$$(D, z) \mapsto Dz$$

of $(\mathfrak{D}, K)$ into K. Each element z of K gives therefore a K-linear functional of $\mathfrak{D}$. This functional is denoted by dz. We have

$$d(yz) = y\,dz + z\,dy,$$

$$d(y + z) = dy + dz.$$

These linear functionals form a subspace $\mathfrak{F}$ of the dual space of $\mathfrak{D}$, if we define $y\,dz$ by $(D, y\,dz) \mapsto yDz$.

Proposition 5.5. *Assume that K is a separably generated and finitely generated extension of k of transcendence degree r. Then the vector space $\mathfrak{D}$ (over K) of derivations of K over k has dimension r. Elements $t_1, \ldots, t_r$ of K from a separating transcendence base of K over k if and only if $dt_1, \ldots, dt_r$ form a basis of the dual space of $\mathfrak{D}$ over K.*

Proof. If $t_1, \ldots, t_r$ is a separating transcendence base for K over k, then we can find derivations $D_1, \ldots, D_r$ of K over k such that $D_i t_j = \delta_{ij}$, by Cases 1 and 2 of the extension theorem. Given $D \in \mathfrak{D}$, let $w_i = Dt_i$. Then clearly $D = \sum w_i D_i$, and so the D_i form a basis for $\mathfrak{D}$ over K, and the dt_i form the dual basis. Conversely, if $dt_1, \ldots, dt_r$ is a basis for $\mathfrak{F}$ over K, and if K is not separably generated over $k(t)$, then by Cases 2 and 3 we can find a derivation D which is trivial on $k(t)$ but nontrivial on K. If $D_1, \ldots, D_r$ is the dual basis of $dt_1, \ldots, dt_r$ (so $D_i t_j = \delta_{ij}$) then $D, D_1, \ldots, D_r$ would be linearly independent over K, contradicting the first part of the theorem.

Corollary 5.6. *Let K be a finitely generated and separably generated extension of k. Let z be an element of K transcendental over k. Then K is separable over $k(z)$ if and only if there exists a derivation D of K over k such that $Dz \neq 0$.*

Proof. If K is separable over $k(z)$, then z can be completed to a separating base of K over k and we can apply the proposition. If $Dz \neq 0$, then $dz \neq 0$, and we can complete dz to a basis of $\mathfrak{F}$ over K. Again from the proposition, it follows that K will be separable over $k(z)$.

Note. Here we have discussed derivations of fields. For derivations in the context of rings and modules, see Chapter XVI.

As an application, we prove:

Theorem 5.7. (Zariski–Matsusaka). *Let K be a finitely generated separable extension of a field k. Let $y, z \in K$ and $z \notin K^p k$ if the characteristic is $p > 0$. Let u be transcendental over K, and put $k_u = k(u)$, $K_u = K(u)$.*

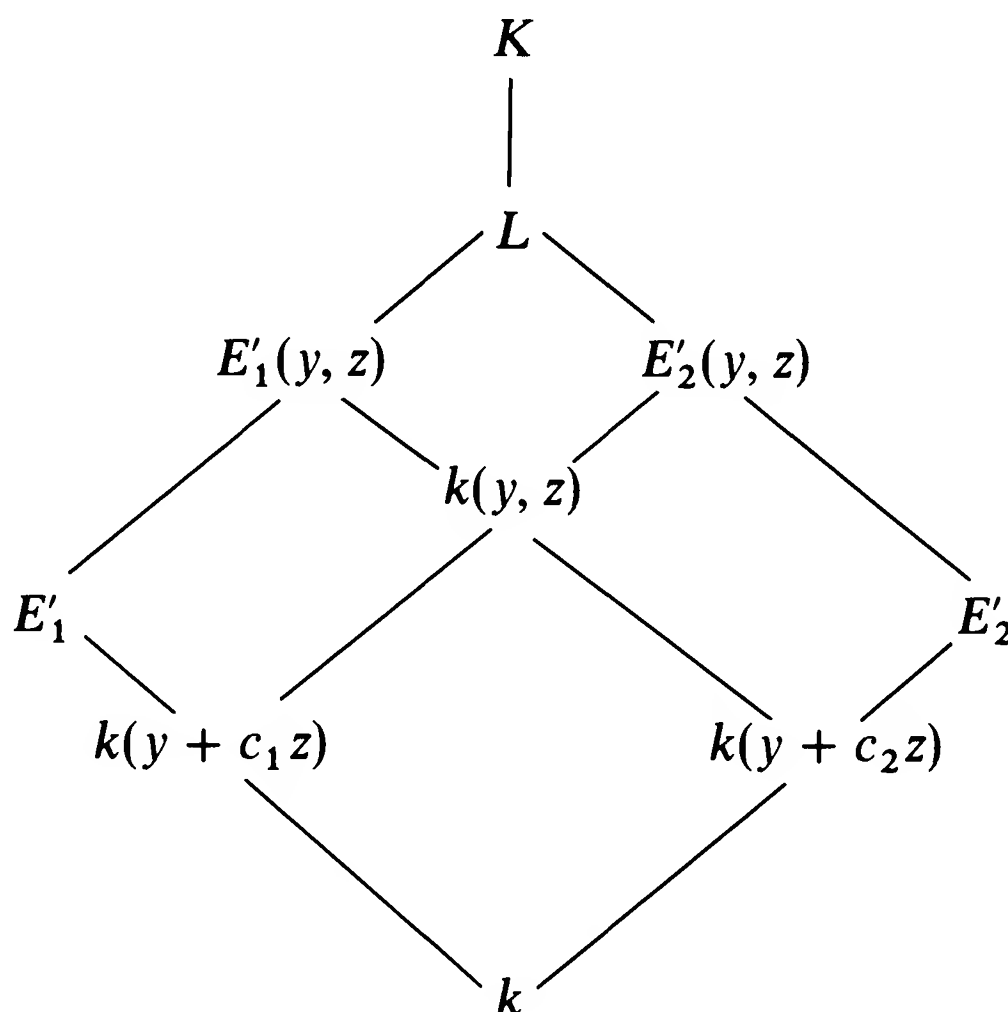
(a) *For all except possibly one value of $c \in k$, K is a separable extension of $k(y + cz)$. Furthermore, K_u is separable over $k_u(y + uz)$.*
(b) *Assume that K is regular over k, and that its transcendence degree is at least 2. Then for all but a finite number of elements $c \in k$, K is a regular extension of $k(y + cz)$. Furthermore, K_u is regular over $k_u(y + uz)$.*

Proof. We shall use throughout the fact that a subfield of a finitely generated extension is also finitely generated (see Exercise 4).

If w is an element of K, and if there exists a derivation D of K over k such that $Dw \neq 0$, then K is separable over $k(w)$, by Corollary 5.6. Also by Corollary 5.6, there exists D such that $Dz \neq 0$. Then for all elements $c \in k$, except possibly one, we have $D(y + cz) = Dy + cDz \neq 0$. Also we may extend D to K_u over k_u by putting $Du = 0$, and then one sees that

$D(y+uz) = Dy + uDz \neq 0$, so K is separable over $k(y + cz)$ except possibly for one value of c, and K_u is separable over $k_u(y + uz)$. In what follows, we assume that the constants $c_1, c_2, \ldots$ are different from the exceptional constant, and hence that K is separable over $k(y + c_i z)$ for $i = 1, 2$.

Assume next that K is regular over k and that the transcendence degree is at least 2. Let $E_i = k(y + c_i z)$ $(i = 1, 2)$ and let E_i' be the algebraic closure of E_i in K. We must show that $E_i' = E_i$ for all but a finite number of constants. Note that $k(y, z) = E_1 E_2$ is the compositum of E_1 and E_2, and that $k(y, z)$ has transcendence degree 2 over k. Hence E_1' and E_2' are free over k. Being subfields of a regular extension of k, they are regular over k, and are therefore linearly disjoint by Theorem 4.12.



By construction, E_1' and E_2' are finite separable algebraic extensions of E_1 and E_2 respectively. Let L be the separable algebraic closure of $k(y, z)$ in K. There is only a finite number of intermediate fields between $k(y, z)$ and L. Furthermore, by Proposition 3.1 the fields $E_1'(y, z)$ and $E_2'(y, z)$ are linearly disjoint over $k(y, z)$. Let c_1 range over the finite number of constants which will exhaust the intermediate extensions between L and $k(y, z)$ obtainable by lifting over $k(y, z)$ a field of type E_i'. If c_2 is now chosen different from any one of these constants c_1, then the only way in which the condition of linear disjointness mentioned above can be compatible with our choice of c_2 is that $E_2'(y, z) = k(y, z)$, i.e. that $E_2' = k(y + c_2 z)$. This means that $k(y + c_2 z)$ is algebraically closed in K, and hence that K is regular over $k(y + c_2 z)$.

As for K_u, let $u_1, u_2, \ldots$ be infinitely many elements algebraically independent over K. Let $k' = k(u_1, u_2, \ldots)$ and $K' = K(u_1, u_2, \ldots)$ be the fields obtained by adjoining these elements to k and K respectively. By what has already been proved, we know that K' is regular over $k'(u + u_i z)$ for all but a finite number of integers i, say for $i = 1$. Our assertion (a) is then a consequence of Corollary 4.14. This concludes the proof of Theorem 5.7.

Theorem 5.8. *Let $K = k(x_1, \dots, x_n) = k(x)$ be a finitely generated regular extension of a field k. Let $u_1, \dots, u_n$ be algebraically independent over $k(x)$. Let*

$$u_{n+1} = u_1 x_1 + \cdots + u_n x_n,$$

and let $k_u = k(u_1, \dots, u_n, u_{n+1})$. Then $k_u(x)$ is separable over k_u, and if the transcendence degree of $k(x)$ over k is $\geqq 2$, then $k_u(x)$ is regular over k_u.

Proof. By the separability of $k(x)$ over k, some x_i does not lie in $K^p k$, say $x_n \notin K^p k$. Then we take

$$y = u_1 x_1 + \cdots + u_{n-1} x_{n-1} \qquad \text{and} \qquad z = x_n,$$

so that $u_{n+1} = y + u_n z$, and we apply Theorem 5.7 to conclude the proof.

Remark. In the geometric language of the next chapter, Theorem 5.8 asserts that the intersection of a k-variety with a generic hyperplane

$$u_1 X_1 + \cdots + u_n X_n - u_{n+1} = 0$$

is a k_u-variety, if the dimension of the k-variety is $\geqq 2$. In any case, the extension $k_u(x)$ is separable over k_u.

EXERCISES

1. Prove that the complex numbers have infinitely many automorphisms. [*Hint*: Use transcendence bases.] Describe all automorphisms and their cardinality.

2. A subfield k of a field K is said to be algebraically closed in K if every element of K which is algebraic over k is contained in k. Prove: If k is algebraically closed in K, and K, L are free over k, and L is separable over k or K is separable over k, then L is algebraically closed in KL.

3. Let $k \subset E \subset K$ be extension fields. Show that

$$\text{tr. deg. } (K/k) = \text{tr. deg. } (K/E) + \text{tr. deg. } (E/k).$$

If $\{x_i\}$ is a transcendence base of E/k, and $\{y_j\}$ is a transcendence base of K/E, then $\{x_i, y_j\}$ is a transcendence base of K/k.

4. Let K/k be a finitely generated extension, and let $K \supset E \supset k$ be a subextension. Show that E/k is finitely generated.

5. Let k be a field and $k(x_1, \dots, x_n) = k(x)$ a finite separable extension. Let $u_1, \dots, u_n$ be algebraically independent over k. Let

$$w = u_1 x_1 + \cdots + u_n x_n.$$

Let $k_u = k(u_1, \dots, u_n)$. Show that $k_u(w) = k_u(x)$.

6. Let $k(x) = k(x_1, \dots, x_n)$ be a separable extension of transcendence degree $r \geqq 1$. Let u_{ij} ($i = 1, \dots, r$; $j = 1, \dots, n$) be algebraically independent over $k(x)$. Let

$$y_i = \sum_{j=1}^{n} u_{ij} x_j.$$

Let $k_u = k(u_{ij})_{\text{all } i, j}$.
(a) Show that $k_u(x)$ is separable algebraic over $k(y_1, \dots, y_r)$.
(b) Show that there exists a polynomial $P(u) \in k[u]$ having the following property. Let $(c) = (c_{ij})$ be elements of k such that $P(c) \neq 0$. Let

$$y'_i = \sum_{j=1}^{n} c_{ij} x_j.$$

Then $k(x)$ is separable algebraic over $k(y')$.

7. Let k be a field and $k[x_1, \dots, x_n] = R$ a finitely generated entire ring over k with quotient field $k(x)$. Let L be a finite extension of $k(x)$. Let I be the integral closure of R in L. Show that I is a finite R-module. [Use Noether normalization, and deal with the inseparability problem and the separable case in two steps.]

8. Let D be a derivation of a field K. Then $D^n: K \to K$ is a linear map. Let $P_n = \operatorname{Ker} D^n$, so P_n is an additive subgroup of K. An element $x \in K$ is called a **logarithmic derivative** (in K) if there exists $y \in K$ such that $x = Dy/y$. Prove:
(a) An element $x \in K$ is the logarithmic derivative of an element $y \in P_n$ but $y \notin P_{n-1}$ ($n > 0$) if and only if

$$(D + x)^n(1) = 0 \qquad \text{and} \qquad (D + x)^{n-1}(1) \neq 0.$$

(b) Assume that $K = \bigcup P_n$, i.e. given $x \in K$ then $x \in P_n$ for some $n > 0$. Let F be a subfield of K such that $DF \subset F$. Prove that x is a logarithmic derivative in F if and only if x is a logarithmic derivative in K. [*Hint*: If $x = Dy/y$ then $(D + x) = y^{-1} D \circ y$ and conversely.]

9. Let k be a field of characteristic 0, and let $z_1, \dots, z_r$ be algebraically independent over k. Let (e_{ij}), $i = 1, \dots, m$ and $j = 1, \dots, r$ be a matrix of integers with $r \geqq m$, and assume that this matrix has rank m. Let

$$w_i = z_1^{e_{i1}} \cdots z_r^{e_{ir}} \qquad \text{for} \quad i = 1, \dots, m.$$

Show that $w_1, \dots, w_m$ are algebraically independent over k. [*Hint*: Consider the K-homomorphism mapping the K-space of derivations of K/k into $K^{(r)}$ given by

$$D \mapsto (Dz_1/z_1, \dots, Dz_r/z_r),$$

and derive a linear condition for those D vanishing on $k(w_1, \dots, w_m)$.]

10. Let k, (z) be as in Exercise 9. Show that if P is a rational function then

$$d(P(z)) = \operatorname{grad} P(z) \cdot dz,$$

using vector notation, i.e. $dz = (dz_1, \dots, dz_r)$ and $\operatorname{grad} P = (D_1 P, \dots, D_r P)$. Define $d \log P$ and express it in terms of coordinates. If P, Q are rational functions in $k(z)$ show that

$$d \log(PQ) = d \log P + d \log Q.$$

CHAPTER IX

Algebraic Spaces

This chapter gives the basic results concerning solutions of polynomial equations in several variables over a field k. First it will be proved that if such equations have a common zero in some field, then they have a common zero in the algebraic closure of k, and such a zero can be obtained by the process known as specialization. However, it is useful to deal with transcendental extensions of k as well. Indeed, if $\mathfrak{p}$ is a prime ideal in $k[X] = k[X_1, \ldots, X_n]$, then $k[X]/\mathfrak{p}$ is a finitely generated ring over k, and the images x_i of X_i in this ring may be transcendental over k, so we are led to consider such rings.

Even if we want to deal only with polynomial equations over a field, we are led in a natural way to deal with equations over the integers $\mathbf{Z}$. Indeed, if the equations are homogeneous in the variables, then we shall prove in §3 and §4 that there are universal polynomials in their coefficients which determine whether these equations have a common zero or not. "Universal" means that the coefficients are integers, and any given special case comes from specializing these universal polynomials to the special case.

Being led to consider polynomial equations over $\mathbf{Z}$, we then consider ideals $\mathfrak{a}$ in $\mathbf{Z}[X]$. The zeros of such an ideal form what is called an algebraic space. If $\mathfrak{p}$ is a prime ideal, the zeros of $\mathfrak{p}$ form what is called an arithmetic variety. We shall meet the first example in the discussion of elimination theory, for which I follow van der Waerden's treatment in the first two editions of his *Moderne Algebra*, Chapter XI.

However, when taking the polynomial ring $\mathbf{Z}[X]/\mathfrak{a}$ for some ideal $\mathfrak{a}$, it usually happens that such a factor ring has divisors of zero, or even nilpotent elements. Thus it is also natural to consider arbitrary commutative rings, and to lay the foundations of algebraic geometry over arbitrary commutative rings as did Grothendieck. We give some basic definitions for this purpose in §5. Whereas the present chapter gives the flavor of algebraic geometry dealing with specific polynomial ideals, the next chapter gives the flavor of geometry developing from commutative algebra, and its systematic application to the more general cases just mentioned.

The present chapter and the next will also serve the purpose of giving the reader an introduction to books on algebraic geometry, notably Hartshorne's systematic basic account. For instance, I have included those results which are needed for Hartshorne's Chapter I and II.

§1. HILBERT'S NULLSTELLENSATZ

The Nullstellensatz has to do with a special case of the extension theorem for homomorphisms, applied to finitely generated rings over fields.

Theorem 1.1. *Let k be a field, and let $k[x] = k[x_1, \dots, x_n]$ be a finitely generated ring over k. Let $\varphi : k \to L$ be an embedding of k into an algebraically closed field L. Then there exists an extension of φ to a homomorphism of $k[x]$ into L.*

Proof. Let $\mathfrak{M}$ be a maximal ideal of $k[x]$. Let σ be the canonical homomorphism $\sigma : k[x] \to k[x]/\mathfrak{M}$. Then $\sigma k[\sigma x_1, \dots, \sigma x_n]$ is a field, and is in fact an extension field of σk. If we can prove our theorem when the finitely generated ring is in fact a field, then we apply $\varphi \circ \sigma^{-1}$ on σk and extend this to a homomorphism of $\sigma k[\sigma x_1, \dots, \sigma x_n]$ into L to get what we want.

Without loss of generality, we therefore assume that $k[x]$ is a field. If it is algebraic over k, we are done (by the known result for algebraic extensions). Otherwise, let $t_1, \dots, t_r$ be a transcendence basis, $r \geqq 1$. Without loss of generality, we may assume that φ is the identity on k. Each element $x_1, \dots, x_n$ is algebraic over $k(t_1, \dots, t_r)$. If we multiply the irreducible polynomial $\mathrm{Irr}(x_i, k(t), X)$ by a suitable non-zero element of $k[t]$, then we get a polynomial all of whose coefficients lie in $k[t]$. Let $a_1(t), \dots, a_n(t)$ be the set of the leading coefficients of these polynomials, and let $a(t)$ be their product,

$$a(t) = a_1(t) \cdots a_n(t).$$

Since $a(t) \neq 0$, there exist elements $t'_1, \dots, t'_r \in k^{\mathrm{a}}$ such that $a(t') \neq 0$, and hence $a_i(t') \neq 0$ for any i. Each x_i is integral over the ring

$$k\left[t_1, \dots, t_r, \frac{1}{a_1(t)}, \dots, \frac{1}{a_r(t)}\right].$$

Consider the homomorphism

$$\varphi : k[t_1, \dots, t_r] \to k^{\mathrm{a}}$$

such that φ is the identity on k, and $\varphi(t_j) = t'_j$. Let $\mathfrak{p}$ be its kernel. Then $a(t) \notin \mathfrak{p}$.

Our homomorphism φ extends uniquely to the local ring $k[t]_{\mathfrak{p}}$ and by the preceding remarks, it extends to a homomorphism of

$$k[t]_{\mathfrak{p}}[x_1, \dots, x_n]$$

into k^a, using Proposition 3.1 of Chapter VII. This proves what we wanted.

Corollary 1.2. *Let k be a field and $k[x_1, \dots, x_n]$ a finitely generated extension ring of k. If $k[x]$ is a field, then $k[x]$ is algebraic over k.*

Proof. All homomorphisms of a field are isomorphisms (onto the image), and there exists a homomorphism of $k[x]$ over k into the algebraic closure of k.

Corollary 1.3. *Let $k[x_1, \dots, x_n]$ be a finitely generated entire ring over a field k, and let $y_1, \dots, y_m$ be non-zero elements of this ring. Then there exists a homomorphism*

$$\psi : k[x] \to k^a$$

over k such that $\psi(y_j) \neq 0$ for all $j = 1, \dots, m$.

Proof. Consider the ring $k[x_1, \dots, x_n, y_1^{-1}, \dots, y_m^{-1}]$ and apply the theorem to this ring.

Let S be a set of polynomials in the polynomial ring $k[X_1, \dots, X_n]$ in n variables. Let L be an extension field of k. By a **zero** of S in L one means an n-tuple of elements $(c_1, \dots, c_n)$ in L such that

$$f(c_1, \dots, c_n) = 0$$

for all $f \in S$. If S consists of one polynomial f, then we also say that (c) is a zero of f. The set of all zeros of S is called an **algebraic set** in L (or more accurately in $L^{(n)}$). Let $\mathfrak{a}$ be the ideal generated by all elements of S. Since $S \subset \mathfrak{a}$ it is clear that every zero of $\mathfrak{a}$ is also a zero of S. However, the converse obviously holds, namely every zero of S is also a zero of $\mathfrak{a}$ because every element of $\mathfrak{a}$ is of type

$$g_1(X)f_1(X) + \cdots + g_m(X)f_m(X)$$

with $f_j \in S$ and $g_i \in k[X]$. Thus when considering zeros of a set S, we may just consider zeros of an ideal. We note parenthetically that every ideal is finitely generated, and so every algebraic set is the set of zeros of a finite number of polynomials. As another corollary of Theorem 1.1, we get:

Theorem 1.4. *Let $\mathfrak{a}$ be an ideal in $k[X] = k[X_1, \dots, X_n]$. Then either $\mathfrak{a} = k[X]$ or $\mathfrak{a}$ has a zero in k^a.*

Proof. Suppose $\mathfrak{a} \neq k[X]$. Then $\mathfrak{a}$ is contained in some maximal ideal $\mathfrak{m}$, and $k[X]/\mathfrak{m}$ is a field, which is a finitely generated extension of k, because it is generated by the images of $X_1, \dots, X_n \bmod \mathfrak{m}$. By Corollary 2.2, this field is algebraic over k, and can therefore be embedded in the algebraic closure k^a. The homomorphism on $k[X]$ obtained by the composition of the canonical map mod $\mathfrak{m}$, followed by this embedded gives the desired zero of $\mathfrak{a}$, and concludes the proof of the theorem.

In §3 we shall consider conditions on a family of polynomials to have a common zero. Theorem 1.4 implies that if they have a common zero in some field, then they have a common zero in the algebraic closure of the field generated by their coefficients over the prime field.

Theorem 1.5. (Hilbert's Nullstellensatz). *Let $\mathfrak{a}$ be an ideal in $k[X]$. Let f be a polynomial in $k[X]$ such that $f(c) = 0$ for every zero $(c) = (c_1, \dots, c_n)$ of $\mathfrak{a}$ in k^a. Then there exists an integer $m > 0$ such that $f^m \in \mathfrak{a}$.*

Proof. We may assume that $f \neq 0$. We use the Rabinowitsch trick of introducing a new variable Y, and of considering the ideal $\mathfrak{a}'$ generated by $\mathfrak{a}$ and $1 - Yf$ in $k[X, Y]$. By Theorem 1.4, and the current assumption, the ideal $\mathfrak{a}'$ must be the whole polynomial ring $k[X, Y]$, so there exist polynomials $g_i \in k[X, Y]$ and $h_i \in \mathfrak{a}$ such that

$$1 = g_0(1 - Yf) + g_1 h_1 + \cdots + g_r h_r.$$

We substitute f^{-1} for Y and multiply by an appropriate power f^m of f to clear denominators on the right-hand side. This concludes the proof.

For questions involving how effective the Nullstellensatz can be made, see the following references also related to the discussion of elimination theory discussed later in this chapter.

Bibliography

[BeY 91] C. Berenstein and A. Yger, Effective Bezout identities in $\mathbf{Q}[z_1, \dots, z_n]$, *Acta Math.* **166** (1991), pp. 69–120

[Br 87] D. Brownawell, Bounds for the degree in Nullstellensatz, *Ann. of Math.* **126** (1987), pp. 577–592

[Br 88] D. Brownawell, Local diophantine nullstellen inequalities, *J. Amer. Math. Soc.* **1** (1988), pp. 311–322

[Br 89] D. Brownawell, Applications of Cayley-Chow forms, *Springer Lecture Notes* **1380:** *Number Theory, Ulm 1987*, H. P. Schlickewei and E. Wirsing (eds.), pp. 1–18

[Ko 88] J. Kollar, Sharp effective nullstellensatz, *J. Amer. Math. Soc.* **1 No. 4** (1988), pp. 963–975

§2. ALGEBRAIC SETS, SPACES AND VARIETIES

We shall make some very elementary remarks on algebraic sets. Let k be a field, and let A be an algebraic set of zeros in some fixed algebraically closed extension field of k. The set of all polynomials $f \in k[X_1, \dots, X_n]$ such that $f(x) = 0$ for all $(x) \in A$ is obviously an ideal $\mathfrak{a}$ in $k[X]$, and is determined by A. We shall call it the ideal **belonging** to A, or say that it is **associated** with A. If A is the set of zeros of a set S of polynomials, then $S \subset \mathfrak{a}$, but $\mathfrak{a}$ may be bigger than S. On the other hand, we observe that A is also the set of zeros of $\mathfrak{a}$.

Let A, B be algebraic sets, and $\mathfrak{a}$, $\mathfrak{b}$ their associated ideals. Then it is clear that $A \subset B$ if and only if $\mathfrak{a} \supset \mathfrak{b}$. Hence $A = B$ if and only if $\mathfrak{a} = \mathfrak{b}$. This has an important consequence. Since the polynomial ring $k[X]$ is Noetherian, it follows that algebraic sets satisfy the dual property, namely every descending sequence of algebraic sets

$$A_1 \supset A_2 \supset \cdots$$

must be such that $A_m = A_{m+1} = \cdots$ for some integer m, i.e. all A_ν are equal for $\nu \geqq m$. Furthermore, dually to another property characterizing the Noetherian condition, we conclude that every non-empty set of algebraic sets contains a minimal element.

Theorem 2.1. *The finite union and the finite intersection of algebraic sets are algebraic sets. If A, B are the algebraic sets of zeros of ideals $\mathfrak{a}$, $\mathfrak{b}$, respectively, then $A \cup B$ is the set of zeros of $\mathfrak{a} \cap \mathfrak{b}$ and $A \cap B$ is the set of zeros of* $(\mathfrak{a}, \mathfrak{b})$.

Proof. We first consider $A \cup B$. Let $(x) \in A \cup B$. Then (x) is a zero of $\mathfrak{a} \cap \mathfrak{b}$. Conversely, let (x) be a zero of $\mathfrak{a} \cap \mathfrak{b}$, and suppose $(x) \notin A$. There exists a polynomial $f \in \mathfrak{a}$ such that $f(x) \neq 0$. But $\mathfrak{a}\mathfrak{b} \subset \mathfrak{a} \cap \mathfrak{b}$ and hence $(fg)(x) = 0$ for all $g \in \mathfrak{b}$, whence $g(x) = 0$ for all $g \in \mathfrak{b}$. Hence (x) lies in B, and $A \cup B$ is an algebraic set of zeros of $\mathfrak{a} \cap \mathfrak{b}$.

To prove that $A \cap B$ is an algebraic set, let $(x) \in A \cap B$. Then (x) is a zero of $(\mathfrak{a}, \mathfrak{b})$. Conversely, let (x) be a zero of $(\mathfrak{a}, \mathfrak{b})$. Then obviously $(x) \in A \cap B$, as desired. This proves our theorem.

An algebraic set V is called **k-irreducible** if it cannot be expressed as a union $V = A \cup B$ of algebraic sets A, B with A, B distinct from V. We also say irreducible instead of **k-irreducible**.

Theorem 2.2. *Let A be an algebraic set.*

(i) *Then A can be expressed as a finite union of irreducible algebraic sets $A = V_1 \cup \ldots \cup V_r$.*

(ii) *If there is no inclusion relation among the V_i, i.e. if $V_i \not\subset V_j$ for $i \neq j$, then the representation is unique.*

(iii) *Let $W, V_1, \dots, V_r$ be irreducible algebraic sets such that*

$$W \subset V_1 \cup \dots \cup V_r.$$

Then $W \subset V_i$ for some i.

Proof. We first show existence. Suppose the set of algebraic sets which cannot be represented as a finite union of irreducible ones is not empty. Let V be a minimal element in its. Then V cannot be irreducible, and we can write $V = A \cup B$ where A, B are algebraic sets, but $A \neq V$ and $B \neq V$. Since each one of A, B is strictly smaller than V, we can express A, B as finite unions of irreducible algebraic sets, and thus get an expression for V, contradiction.

The uniqueness will follow from (iii), which we prove next. Let W be contained in the union $V_1 \cup \dots \cup V_r$. Then

$$W = (W \cap V_1) \cup \dots \cup (W \cap V_r).$$

Since each $W \cap V_i$ is an algebraic set, by the irreducibility of W we must have $W = W \cap V_i$ for some i. Hence $W \subset V_i$ for some i, thus proving (iii).

Now to prove (ii), apply (iii) to each W_j. Then for each j there is some i such that $W_j \subset V_i$. Similarly for each i there exists ν such that $V_i \subset W_\nu$. Since there is no inclusion relation among the W_j's, we must have $W_j = V_i = W_\nu$. This proves that each W_j appears among the V_i's and each V_i appears among the W_j's, and proves the uniqueness of the representation. It also concludes the proof of Theorem 2.2.

Theorem 2.3 *An algebraic set is irreducible if and only if its associated ideal is prime.*

Proof. Let V be irreducible and let $\mathfrak{p}$ be its associated ideal. If $\mathfrak{p}$ is not prime, we can find two polynomials f, $g \in k[X]$ such that $f \notin \mathfrak{p}$, $g \notin \mathfrak{p}$, but $fg \in \mathfrak{p}$. Let $\mathfrak{a} = (\mathfrak{p}, f)$ and $\mathfrak{b} = (\mathfrak{p}, g)$. Let A be the algebraic set of zeros of $\mathfrak{a}$, and B the algebraic set of zeros of $\mathfrak{b}$. Then $A \subset V$, $A \neq V$ and $B \subset V$, $B \neq V$. Furthermore $A \cup B = V$. Indeed, $A \cup B \subset V$ trivially. Conversely, let $(x) \in V$. Then $(fg)(x) = 0$ implies $f(x)$ or $g(x) = 0$. Hence $(x) \in A$ or $(x) \in B$, proving $V = A \cup B$, and V is not irreducible. Conversely, let V be the algebraic set of zeros of a prime ideal $\mathfrak{p}$. Suppose $V = A \cup B$ with $A \neq V$ and $B \neq V$. Let $\mathfrak{a}$, $\mathfrak{b}$ be the ideals associated with A and B respectively. There exist polynomials $f \in \mathfrak{a}$, $f \notin \mathfrak{p}$ and $g \in \mathfrak{b}$, $g \notin \mathfrak{p}$. But fg vanishes on $A \cup B$ and hence lies in $\mathfrak{p}$, contradiction which proves the theorem.

Warning. Given a field k and a prime ideal $\mathfrak{p}$ in $k[X]$, it may be that the ideal generated by $\mathfrak{p}$ in $k^{\mathrm{a}}[X]$ is not prime, and the algebraic set defined over k^{a} by $\mathfrak{p}k^{\mathrm{a}}[X]$ has more than one component, and so is not irreducible. Hence the prefix referring to k is really necessary.

It is also useful to extend the terminology of algebraic sets as follows. Given an ideal $\mathfrak{a} \subset k[X]$, to each field K containing k we can associate to $\mathfrak{a}$ the set

$\mathfrak{X}_{\mathfrak{a}}(K)$ consisting of the zeros of $\mathfrak{a}$ in K. Thus $\mathfrak{X}_{\mathfrak{a}}$ is an association

$$\mathfrak{X}_{\mathfrak{a}}: K \mapsto \mathfrak{X}_{\mathfrak{a}}(K) \subset K^{(n)}.$$

We shall speak of $\mathfrak{X}_{\mathfrak{a}}$ itself as an **algebraic space**, so that $\mathfrak{X}_{\mathfrak{a}}$ is not a set, but to each field K associates the set $\mathfrak{X}_{\mathfrak{a}}(K)$. Thus $\mathfrak{X}_{\mathfrak{a}}$ is a functor from extensions K of k to sets (functorial with respect to field isomorphisms). By a k-**variety** we mean the algebraic space associated with a prime ideal $\mathfrak{p}$.

The notion of associated ideal applies also to such $\mathfrak{X}_{\mathfrak{a}}$, and the associated ideal of $\mathfrak{X}_{\mathfrak{a}}$ is also rad($\mathfrak{a}$). We shall omit the subscript $\mathfrak{a}$ and write simply $\mathfrak{X}$ for this generalized notion of algebraic space. Of course we have

$$\mathfrak{X}_{\mathfrak{a}} = \mathfrak{X}_{\mathrm{rad}(\mathfrak{a})}.$$

We say that $\mathfrak{X}_{\mathfrak{a}}(K)$ is the set of **points of** $\mathfrak{X}_{\mathfrak{a}}$ **in** K. By the Hilbert Nullstellensatz, Theorem 1.1, it follows that if $K \subset K'$ are two algebraically closed fields containing k, then the ideals associated with $\mathfrak{X}_{\mathfrak{a}}(K)$ and $\mathfrak{X}_{\mathfrak{a}}(K')$ are equal to each other, and also equal to rad($\mathfrak{a}$). Thus the smallest algebraically closed field k^{a} containing k already determines these ideals. However, it is also useful to consider larger fields which contain transcendental elements, as we shall see.

As another example, consider the polynomial ring $k[X_1, \ldots, X_n] = k[X]$. Let $\mathbf{A}^n$ denote the algebraic space associated with the zero ideal. Then $\mathbf{A}^n$ is called **affine** n-space. Let K be a field containing k. For each n-tuple $(c_1, \ldots, c_n) \in K^{(n)}$ we get a homomorphism

$$\varphi: k[X_1, \ldots, X_n] \to K$$

such that $\varphi(X_i) = c_i$ for all i. Thus points in $\mathbf{A}^n(K)$ correspond bijectively to homomorphisms of $k(X)$ into K.

More generally, let V be a k-variety with associated prime ideal $\mathfrak{p}$. Then $k[X]/\mathfrak{p}$ is entire. Denote by ξ_i the image of X_i under the canonical homomorphism $k[X] \to k[X]/\mathfrak{p}$. We call (ξ) **the generic point** of V over k. On the other hand, let (x) be a point of V in some field K. Then $\mathfrak{p}$ vanishes on (x), so the homomorphism $\varphi : k[X] \to k[x]$ sending $X_i \mapsto x_i$ factors through $k[X]/\mathfrak{p} = k[\xi]$, whence we obtain a natural homomorphism $k[\xi] \to k[x]$. If this homomorphism is an isomorphism, then we call (x) **a generic point** of V in K.

Given two points $(x) \in \mathbf{A}^n(K)$ and $(x') \in \mathbf{A}^n(K')$, we say that (x') is a **specialization** of (x) (over k) if the map $x_i \mapsto x_i'$ is induced by a homomorphism $k[x] \to k[x']$. From the definition of a generic point of a variety, it is then immediate that:

A variety V is the set of specializations of its generic point, or of a generic point.

In other words, $V(K)$ is the set of specializations of (ξ) in K for every field K containing k.

Let us look at the converse construction of algebraic sets. Let $(x) = (x_1, \ldots, x_n)$ be an n-tuple with coordinates $x_i \in K$ for some extension field K of k. Let $\mathfrak{p}$ be the ideal in $k[X]$ consisting of all polynomials $f(X)$ such that

$f(x) = 0$. We call $\mathfrak{p}$ the ideal **vanishing** on (x). Then $\mathfrak{p}$ is prime, because if $fg \in \mathfrak{p}$ so $f(x)g(x) = 0$, then $f \in \mathfrak{p}$ or $g \in \mathfrak{p}$ since K has no divisors of 0. Hence $\mathfrak{X}_{\mathfrak{p}}$ is a k-variety V, and (x) is a generic point of V over k because $k[X]/\mathfrak{p} \approx k[x]$.

For future use, we state the next result for the polynomial ring over a factorial ring rather than over a field.

Theorem 2.4. *Let R be a factorial ring, and let $W_1, \ldots, W_m$ be m independent variables over its quotient field k. Let $k(w_1, \ldots, w_m)$ be an extension of transcendence degree $m - 1$. Then the ideal in $R[W]$ vanishing on (w) is principal.*

Proof. By hypothesis there is some polynomial $P(W) \in R[W]$ of degree $\geqq 1$ vanishing on (w), and after taking an irreducible factor we may assume that this polynomial is irreducible, and so is a prime element in the factorial ring $R[W]$. Let $G(W) \in R[W]$ vanish on (w). To prove that P divides G, after selecting some irreducible factor of G vanishing on (w) if necessary, we may assume without loss of generality that G is a prime element in $R[W]$. One of the variables W_i occurs in $P(W)$, say W_m, so that w_m is algebraic over $k(w_1, \ldots, w_{m-1})$. Then $(w_1, \ldots, w_{m-1})$ are algebraically independent, and hence W_m also occurs in G. Furthermore, $P(w_1, \ldots, w_{m-1}, W_m)$ is irreducible as a polynomial in $k(w_1, \ldots, w_{m-1})[W_m]$ by the Gauss lemma as in Chapter IV, Theorem 2.3. Hence there exists a polynomial $H(W_m) \in k(w_1, \ldots, w_{m-1})[W_m]$ such that

$$G(W) = H(W_m)P(W).$$

Let $R' = R[w_1, \ldots, w_{m-1}]$. Then P, G have content 1 as polynomials in $R'[W_m]$. By Chapter IV Corollary 2.2 we conclude that $H \in R'[W_m] \approx R[W]$, which proves Theorem 2.4.

Next we consider homogeneous ideals and projective space. A polynomial $f(X) \in k[X]$ can be written as a linear combination

$$f(X) = \sum c_{(\nu)} M_{(\nu)}(X)$$

with monomials $M_{(\nu)}(X) = X_1^{\nu_1} \cdots X_n^{\nu_n}$ and $c_{(\nu)} \in k$. We denote the **degree** of $M_{(\nu)}$ by

$$|\nu| = \deg M_{(\nu)} = \sum \nu_i.$$

If in this expression for f the degrees of the monomials $X^{(\nu)}$ are all the same (whenever the coefficient $c_{(\nu)}$ is $\neq 0$), then we say that f is a **form**, or also that f is a **homogeneous** (of that degree). An arbitrary polynomial $f(X)$ in $K[X]$ can also be written

$$f(X) = \sum f^{(d)}(X),$$

where each $f^{(d)}$ is a form of degree d (which may be 0). We call $f^{(d)}$ the **homogeneous part** of f of degree d.

An ideal $\mathfrak{a}$ of $k[X]$ is called **homogeneous** if whenever $f \in \mathfrak{a}$ then each homogeneous part $f^{(d)}$ also lies in $\mathfrak{a}$.

Proposition 2.5. *An ideal $\mathfrak{a}$ is homogeneous if and only if $\mathfrak{a}$ has a set of generators over $k[X]$ consisting of forms.*

Proof. Suppose $\mathfrak{a}$ is homogeneous and that $f_1, \ldots, f_r$ are generators. By hypothesis, for each integer $d \geqq 0$ the homogeneous components $f_i^{(d)}$ also lie in $\mathfrak{a}$, and the set of such $f_i^{(d)}$ (for all i, d) form a set of homogeneous generators. Conversely, let f be a homogeneous element in $\mathfrak{a}$ and let $g \in K[X]$ be arbitrary. For each d, $g^{(d)}f$ lies in $\mathfrak{a}$, and $g^{(d)}f$ is homogeneous, so all the homogeneous components of gf also lie in $\mathfrak{a}$. Applying this remark to the case when f ranges over a set of homogeneous generators for $\mathfrak{a}$ shows that $\mathfrak{a}$ is homogeneous, and concludes the proof of the proposition.

An algebraic space $\mathfrak{X}$ is called **homogeneous** if for every point $(x) \in \mathfrak{X}$ and t transcencental over $k(x)$, the point (tx) also lies in $\mathfrak{X}$. If t, u are transcendental over $k(x)$, then there is an isomorphism

$$k[x, t] \xrightarrow{\approx} k[x, u]$$

which sends t on u and restricts to the identity on $k[x]$, so to verify the above condition, it suffices to verify it for some transcendental t over $k(x)$.

Proposition 2.6. *An algebraic space $\mathfrak{X}$ is homogeneous if and only if its associated ideal $\mathfrak{a}$ is homogeneous.*

Proof. Suppose $\mathfrak{X}$ is homogeneous. Let $f(X) \in k[X]$ vanish on $\mathfrak{X}$. For each $(x) \in \mathfrak{X}$ and t transcendental over $k(x)$ we have

$$0 = f(x) = f(tx) = \sum_d t^d f^{(d)}(x).$$

Therefore $f^{(d)}(x) = 0$ for all d, whence $f^{(d)} \in \mathfrak{a}$ for all d. Hence $\mathfrak{a}$ is homogeneous. Conversely, suppose $\mathfrak{a}$ homogeneous. By the Hilbert Nullstellensatz, we know that $\mathfrak{X}$ consists of the zeros of $\mathfrak{a}$, and hence consists of the zeros of a set of homogeneous generators for $\mathfrak{a}$. But if f is one of those homogeneous generators of degree d, and (x) is a point of $\mathfrak{X}$, then for t transcendental over $k(x)$ we have

$$0 = f(x) = t^d f(x) = f(tx),$$

so (tx) is also a zero of $\mathfrak{a}$. Hence $\mathfrak{X}$ is homogeneous, thus proving the proposition.

Proposition 2.7. *Let $\mathfrak{X}$ be a homogeneous algebraic space. Then each irreducible component V of $\mathfrak{X}$ is also homogeneous.*

Proof. Let $V = V_1, \ldots, V_r$ be the irreducible components of $\mathfrak{X}$, without inclusion relation. By Remark 3.3 we know that $V_1 \not\subset V_2 \cup \ldots \cup V_r$, so there is a point $(x) \in V_1$ such that $(x) \notin V_1$ for $i = 2, \ldots, r$. By hypothesis, for t transcendental over $k(x)$ it follows that $(tx) \in \mathfrak{X}$ so $(tx) \in V_i$ for some i. Specializing to $t = 1$, we conclude that $(x) \in V_i$, so $i = 1$, which proves that V_1 is homogeneous, as was to be shown.

Let V be a variety defined over k by a prime ideal $\mathfrak{p}$ in $k[X]$. Let (x) be a generic point of V over k. We say that (x) is **homogeneous (over k)** if for t

transcendental over $k(x)$, the point (tx) is also a point of V, or in other words, (tx) is a specialization of (x). If this is the case, then we have an isomorphism

$$k[x_1, \ldots, x_n] \approx k[tx_1, \ldots, tx_n],$$

which is the identity on k and sends x_i on tx_i. It then follows from the preceding propositions that the following conditions are equivalent for a variety V over k:

V is homogeneous.
The prime ideal of V in $k[X]$ is homogeneous.
A generic point of V over k is homogeneous.

A homogeneous ideal always has a zero, namely the origin (0), which will be called the **trivial zero**. We shall want to know when a homogeneous algebraic set has a non-trivial zero (in some algebraically closed field). For this we introduce the terminology of projective space as follows. Let (x) be some point in $\mathbf{A}^n$ and λ an element of some field containing $k(x)$. Then we denote by (λx) the point $(\lambda x_1, \ldots, \lambda x_n)$. Two points $(x), (y) \in \mathbf{A}^n(K)$ for some field K are called equivalent if not all their coordinates are 0, and there exists some element $\lambda \in K$, $\lambda \neq 0$, such that $(\lambda x) = (y)$. The equivalence classes of such points in $\mathbf{A}^n(K)$ are called the points of **projective space** in K. We denote this projective space by $\mathbf{P}^{n-1}$, and the set of points of projective space in K by $\mathbf{P}^{n-1}(K)$. We define an **algebraic space in projective space** to be the non-trivial zeros of a homogeneous ideal, with two zeros identified if they differ by a common non-zero factor.

Algebraic spaces over rings

As we shall see in the next section, it is not sufficient to look only at ideals in $k[X]$ for some field k. Sometimes, even often, one wants to deal with polynomial equations over the integers $\mathbf{Z}$, for several reasons. In the example of the next sections, we shall find universal conditions over $\mathbf{Z}$ on the coefficients of a system of forms so that these forms have a non-trivial common zero. Furthermore, in number theory—diophantine questions—one wants to consider systems of equations with integer coefficients, and to determine solutions of these equations in the integers or in the rational numbers, or solutions obtained by reducing mod p for a prime p. Thus one is led to extend the notions of algebraic space and variety as follows. Even though the applications of the next section will be over $\mathbf{Z}$, we shall now give general definitions over an arbitrary commutative ring R.

Let $f(X) \in R[X] = R[X_1, \ldots, X_n]$ be a polynomial with coefficients in R. Let $R \to A$ be an R-algebra, by which for the rest of this chapter we mean a homomorphism of commutative rings. We obtain a corresponding homomorphism

$$R[X] \to A[X]$$

on the polynomial rings, denoted by $f \mapsto f_A$ whereby the coefficients of f_A are the images of the coefficients of f under the homomorphism $R \to A$. By a **zero** of f in A we mean a zero of f_A in A. Similarly, let S be a set of polynomials in $R[X]$. By a **zero** of S in A we mean a common zero in A of all polynomials $f \in S$. Let $\mathfrak{a}$ be the ideal generated by S in $R[X]$. Then a zero of S in A is also

a zero of $\mathfrak{a}$ in A. We denote the set of zeros of S in A by $\mathfrak{Z}_S(A)$, so that we have

$$\mathfrak{Z}_S(A) = \mathfrak{Z}_{\mathfrak{a}}(A).$$

We call $\mathfrak{Z}_{\mathfrak{a}}(A)$ an **algebraic set** over R. Thus we have an association

$$\mathfrak{Z}_{\mathfrak{a}}: A \mapsto \mathfrak{Z}_{\mathfrak{a}}(A)$$

which to each R-algebra associates the set of zeros of $\mathfrak{a}$ in that algebra. We note that R-algebras form a category, whereby a morphism is a ring homomorphism $\varphi: A \to A'$ making the following diagram commutative:

$$\begin{array}{ccc} & & A \\ & \nearrow & \\ R & & \downarrow \varphi \\ & \searrow & \\ & & A' \end{array}$$

Then it is immediately verified that $\mathfrak{Z}_{\mathfrak{a}}$ is a functor from the category of R-algebras to the category of sets. Again we call $\mathfrak{Z}_{\mathfrak{a}}$ an **algebraic space** over R.

If R is Noetherian, then $R[X]$ is also Noetherian (Chapter IV, Theorem 4.1), and so if $\mathfrak{a}$ is an ideal, then there is always some finite set of polynomials S generating the ideal, so $\mathfrak{Z}_S = \mathfrak{Z}_{\mathfrak{a}}$.

The notion of **radical** of $\mathfrak{a}$ is again defined as the set of polynomials $h \in R[X]$ such that $h^N \in \mathfrak{a}$ for some positive integer N. Then the following statement is immediate:

Suppose that R is entire. Then for every R-algebra $R \to K$ with a field K, we have

$$\mathfrak{Z}_{\mathfrak{a}}(K) = \mathfrak{Z}_{\text{rad}(\mathfrak{a})}(K).$$

We can define **affine space $\mathbf{A}^n$** over R. Its points consist of all n-tuples $(x_1, \ldots, x_n) = (x)$ with x_i in some R-algebra A. Thus $\mathbf{A}^n$ is again an association

$$A \mapsto \mathbf{A}^n(A)$$

from R-algebras to sets of points. Such points are in bijection with homormorphisms

$$R[X] \to A$$

from the polynomial ring over R into A. In the next section we shall limit ourselves to the case when $A = K$ is a field, and we shall consider only the functor $K \mapsto \mathbf{A}^n(K)$ for fields K. Furthermore, we shall deal especially with the case when $R = \mathbf{Z}$, so $\mathbf{Z}$ has a unique homomorphism into a field K. Thus a field K can always be viewed as a $\mathbf{Z}$-algebra.

Suppose finally that R is entire (for simplicity). We can also consider projective space over R. Let $\mathfrak{a}$ be an ideal in $R[X]$. We define $\mathfrak{a}$ to be homogeneous just as before. Then a homogeneous ideal in $R[X]$ can be viewed as defining an algebraic subset in projective space $\mathbf{P}^n(K)$ for each field K (as an R-algebra). If $R = \mathbf{Z}$,

then $\mathfrak{a}$ defines an algebraic subset in $\mathbf{P}^n(K)$ for every field K. Similarly, one can define the notion of a homogeneous algebraic space $\mathfrak{X}$ over R, and over the integers $\mathbf{Z}$ *a fortiori*. Propositions 2.6 and 2.7 and their proofs are also valid in this more general case, viewing $\mathfrak{X} = \mathfrak{X}_{\mathfrak{a}}$ as a functor from fields K to sets $\mathbf{P}^n(K)$.

If $\mathfrak{a}$ is a prime ideal $\mathfrak{p}$, then we call $\mathfrak{X}_{\mathfrak{p}}$ an R-**variety** V. If R is Noetherian, so $R[X]$ is Noetherian, it follows as before that an algebraic space $\mathfrak{X}$ over R is a finite union of R-varieties without inclusion relations. We shall carry this out in §5, in the very general context of commutative rings. Just as we did over a field, we may form the factor ring $\mathbf{Z}[X]/\mathfrak{p}$ and the image (x) of (X) in this factor ring is called a **generic point** of V.

§3. PROJECTIONS AND ELIMINATION

Let $(W) = (W_1, \ldots, W_m)$ and $(X) = (X_1, \ldots, X_n)$ be two sets of independent variables. Then ideals in $k[W, X]$ define algebraic spaces in the product space $\mathbf{A}^{m+n}$. Let $\mathfrak{a}$ be an ideal in $k[W, X]$. Let $\mathfrak{a}_1 = \mathfrak{a} \cap k[W]$. Let $\mathfrak{X}$ be the algebraic space of zeros of $\mathfrak{a}$ and let $\mathfrak{X}_1$ be the algebraic space of zeros of $\mathfrak{a}_1$. We have the projection

$$\mathrm{pr}: \mathfrak{X}^{m+n} \to \mathfrak{X}^m \quad \text{or} \quad \mathrm{pr}: \mathbf{A}^{m+n} \to \mathbf{A}^m$$

which maps a point (w, x) to its first set of coordinates (w). It is clear that $\mathrm{pr}\,\mathfrak{X} \subset \mathfrak{X}_1$. In general it is not true that $\mathrm{pr}\,\mathfrak{X} = \mathfrak{X}_1$. For example, the ideal $\mathfrak{p}$ generated by the single polynomial $W_1^2 - W_2X_1 = 0$ is prime. Its intersection with $k[W_1, W_2]$ is the zero ideal. But it is not true that every point in the affine (W_1, W_2)-space is the projection of a point in the variety $\mathfrak{X}_{\mathfrak{p}}$. For instance, the point $(1, 0)$ is not the projection of any zero of $\mathfrak{p}$. One says in such a case that the projection is **incomplete**. We shall now consider a situation when such a phenomenon does not occur.

In the first place, let $\mathfrak{p}$ be a prime ideal in $k[W, X]$ and let V be its variety of zeros. Let (w, x) be a generic point of V. Let $\mathfrak{p}_1 = \mathfrak{p} \cap k[W]$. Then (w) is a generic point of the variety V_1 which is the algebraic space zeros of $\mathfrak{p}_1$. This is immediate from the canonical injective homomorphism

$$k[W]/\mathfrak{p}_1 \to k[W, X]/\mathfrak{p}.$$

Thus the generic point (w) of V_1 is the projection of the generic point (w, x) of V. The question is whether a special point (w') of V_1 is the projection of a point of V.

In the subsequent applications, we shall consider ideals which are homogeneous only in the X-variables, and similarly algebraic subsets which are homogeneous in the second set of variables in $\mathbf{A}^n$.

An ideal $\mathfrak{a}$ in $k[W, X]$ which is homogeneous in (X) defines an algebraic space in $\mathbf{A}^m \times \mathbf{P}^{n-1}$. If V is an irreducible component of the algebraic set defined by $\mathfrak{a}$, then we may view V as a subvariety of $\mathbf{A}^m \times \mathbf{P}^{n-1}$. Let $\mathfrak{p}$ be the prime ideal associated with V. Then $\mathfrak{p}$ is homogeneous in (X). Let $\mathfrak{p}_1 = \mathfrak{p} \cap k[W]$. We shall see that the situation of an incomplete projection mentioned previously is eliminated when we deal with projective space.

We can also consider the product $\mathbf{A}^m \times \mathbf{P}^n$, defined by the zero ideal over $\mathbf{Z}$. For each field K, the set of points of $\mathbf{A}^m \times \mathbf{P}^n$ in K is $\mathbf{A}^m(K) \times \mathbf{P}^n(K)$. An ideal $\mathfrak{a}$ in $\mathbf{Z}[W, X]$, homogeneous in (X), defines an algebraic space $\mathfrak{X} = \mathfrak{X}_\mathfrak{a}$ in $\mathbf{A}^m \times \mathbf{P}^n$. We may form its projection $\mathfrak{X}_1$ on the first factor. This applies in particular when $\mathfrak{a}$ is a prime ideal $\mathfrak{p}$, in which case we call $\mathfrak{X}_\mathfrak{a}$ an **arithmetic subvariety** of $\mathbf{A}^m \times \mathbf{P}^n$. Its projection V_1 is an arithmetic subvariety of $\mathbf{A}^m$, associated with the prime ideal $\mathfrak{p}_1 = \mathfrak{p} \cap \mathbf{Z}[W]$.

Theorem 3.1. *Let $(W) = (W_1, \dots, W_m)$ and $(X) = (X_1, \dots, X_n)$ be independent families of variables. Let $\mathfrak{p}$ be a prime ideal in $k[W, X]$ (resp. $\mathbf{Z}[W, X]$) and assume $\mathfrak{p}$ is homogeneous in (X). Let V be the corresponding irreducible algebraic space in $\mathbf{A}^m \times \mathbf{P}^{n-1}$. Let $\mathfrak{p}_1 = \mathfrak{p} \cap k[W]$ (resp. $\mathfrak{p} \cap \mathbf{Z}[W]$), and let V_1 be the projection of V on the first factor. Then V_1 is the algebraic space of zeros of $\mathfrak{p}_1$ in $\mathbf{A}^m$.*

Proof. Let V have generic point (w, x). We have to prove that every zero (w') of $\mathfrak{p}_1$ in a field is the projection of some zero (w', x') of $\mathfrak{p}$ such that not all the coordinates of (x') are equal to 0. By assumption, not all the coordinates of (x) are equal to 0, since we viewed V as a subset of $\mathbf{A}^m \times \mathbf{P}^{n-1}$. For definiteness, say we are dealing with the case of a field k. By Chapter VII, Proposition 3.3, the homomorphism $k[w] \to k[w']$ can be extended to a place φ of $k(w, x)$. By Proposition 3.4 of Chapter VII, there is some coordinate x_j such that $\varphi(x_i/x_j) \neq \infty$ for all $i = 1, \dots, n$. We let $x_i' = \varphi(x_i/x_j)$ for all i to conclude the proof. The proof is similar when dealing with algebraic spaces over $\mathbf{Z}$, replacing k by $\mathbf{Z}$.

Remarks. Given the point $(w') \in \mathbf{A}^m$, the point (w', x') in $\mathbf{A}^m \times \mathbf{P}^{n-1}$ may of course not lie in $k(w')$. The coordinates (x') could even be transcendental over $k(x')$. By any one of the forms of the Hilbert Nullstellensatz, say Corollary 1.3 of Theorem 1.1, we do know that (x') could be found algebraic over $k(w')$, however. In light of the various versions of the Nullstellensatz, if a set of forms has a non-trivial common zero in some field, then it has a non-trivial common zero in the algebraic closure of the field generated by the coefficients of the forms over the prime field. In a theorem such as Theorem 1.2 below, the conditions on the coefficients for the forms to have a non-trivial common zero (or a zero in projective space) are therefore also conditions for the forms to have such a zero in that algebraic closure.

We shall apply Theorem 3.1 to show that given a finite family of homogeneous polynomials, the property that they have a non-trivial common zero in some

algebraically closed field can be expressed in terms of a finite number of universal polynomial equations in their coefficients. We make this more precise as follows.

Consider a finite set of forms $(f) = (f_1, \ldots, f_r)$. Let $d_1, \ldots, d_r$ be their degrees. *We assume* $d_i \geqq 1$ *for* $i = 1, \ldots, r$. Each f_i can be written

$$f_i = \sum w_{i,(\nu)} M_{(\nu)}(X) \tag{1}$$

where $M_{(\nu)}(X)$ is a monomial in (X) of degree d_i, and $w_{i,(\nu)}$ is a coefficient. We shall say that (f) has a **non-trivial zero** (x) if $(x) \neq (0)$ and $f_i(x) = 0$ for all i. We let $(w) = (w)_f$ be the point obtained by arranging the coefficients $w_{i,(\nu)}$ of the forms in some definite order, and we consider this point as a point in some affine space $\mathbf{A}^m$, where m is the number of such coefficients. This integer m is determined by the given degrees $d_1, \ldots, d_r$. In other words, given such degrees, the set of all forms $(f) = (f_1, \ldots, f_r)$ with these degrees is in bijection with the points of $\mathbf{A}^m$.

Theorem 3.2. (Fundamental theorem of elimination theory.) *Given degrees $d_1, \ldots, d_r$, the set of all forms $(f_1, \ldots, f_r)$ in n variables having a non-trivial common zero is an algebraic subspace of $\mathbf{A}^m$ over $\mathbf{Z}$.*

Proof. Let $(W) = (W_{i,(\nu)})$ be a family of variables independent of (X). Let $(F) = (F_1, \ldots, F_r)$ be the family of polynomials in $\mathbf{Z}[W, X]$ given by

$$F_i(W, X) = \sum W_{i,(\nu)} M_{(\nu)}(X) \tag{2}$$

where $M_{(\nu)}(X)$ ranges over all monomials in (X) of degree d_i, so $(W) = (W)_F$. We call $F_1, \ldots, F_r$ **generic forms**. Let

$$\mathfrak{a} = \text{ideal in } \mathbf{Z}[W, X] \text{ generated by } F_1, \ldots, F_r.$$

Then $\mathfrak{a}$ is homogeneous in (X). Thus we are in the situation of Theorem 3.1, with $\mathfrak{a}$ defining an algebraic space $\mathfrak{A}$ in $\mathbf{A}^m \times \mathbf{P}^{n-1}$. Note that (w) is a specialization of (W), or, as we also say, (f) is a specialization of (F). As in Theorem 3.1, let $\mathfrak{A}_1$ be the projection of $\mathfrak{A}$ on the first factor. Then directly from the definitions, (f) has a non-trivial zero if and only if $(w)_f$ lies in $\mathfrak{A}_1$, so Theorem 3.2 is a special case of Theorem 3.1.

Corollary 3.3. *Let (f) be a family of n forms in n variables, and assume that $(w)_f$ is a generic point of $\mathbf{A}^m$, i.e. that the coefficients of these forms are algebraically independent. Then (f) does not have a non-trivial zero.*

Proof. There exists a specialization of (f) which has only the trivial zero, namely $f_1' = X_1^{d_1}, \ldots, f_n' = X_n^{d_n}$.

Next we follow van der Waerden in showing that $\mathfrak{A}$ and hence $\mathfrak{A}_1$ are irreducible.

Theorem 3.4. *The algebraic space $\mathfrak{A}_1$ of forms having a non-trivial common zero in Theorem 3.2 is actually a $\mathbf{Z}$-variety, i.e. it is irreducible. The prime ideal*

$\mathfrak{p}$ in $\mathbf{Z}[W, X]$ associated with $\mathfrak{A}$ consists of all polynomials $G(W, X) \in \mathbf{Z}[W, X]$ such that for some index j there is an integer $s \geqq 0$ satisfying

$$(*)_j \qquad X_j^s G(W, X) \equiv 0 \bmod (F_1, \ldots, F_r); \text{ that is, } X_j^s G(W, X) \in \mathfrak{a}.$$

If relation $(*)$ holds for one index j, then it holds for every $j = 1, \ldots, n$. (Of course, the integer s depends on j.)

Proof. We construct a generic point of $\mathfrak{A}$. We select any one of the variables, say X_q, and rewrite the forms F_i as follows:

$$F_i(W, X) = F_i^* + Z_i X_q^{d_i}$$

where F_i^* is the sum of all monomials except the monomial containing $X_q^{d_i}$. The coefficients (W) are thereby split into two families, which we denote by (Y) and (Z), where $(Z) = (Z_1, \ldots, Z_r)$ are the coefficients of $(X_q^{d_1}, \ldots, X_q^{d_r})$ in $(F_1, \ldots, F_r)$, and (Y) is the remaining family of coefficients of $F_1^*, \ldots, F_r^*$. We have $(W) = (Y, Z)$, and we may write the polynomials F_i in the form

$$F_i(W, X) = F_i(Y, Z, X) = F_i^*(Y, X) + Z_i X_q^{d_i}.$$

Corresponding to the variables (Y, X) we choose quantities (y, x) algebraically independent over $\mathbf{Z}$. We let

$$(3) \qquad z_i = -F_i^*(y, x)/x_q^{d_i} = -F_i^*(y, x/x_q).$$

We shall prove that (y, z, x) is a generic point of $\mathfrak{A}$.

From our construction, it is immediately clear that $F_i(y, z, x) = 0$ for all i, and consequently if $G(W, X) \in \mathbf{Z}[W, X]$ satisfies $(*)$, then $G(y, z, x) = 0$.

Conversely, let $G(Y, Z, X) \in \mathbf{Z}[Y, Z, X] = \mathbf{Z}[W, X]$ satisfy $G(y, z, x) = 0$. From Taylor's formula in several variables we obtain

$$\begin{aligned} G(Y, Z, X) &= G(Y, \ldots, -F_i^*/X_q^{d_i} + Z_i + F_i^*/X_q^{d_i}, \ldots, X) \\ &= G(Y, -F_i^*/X_q^{d_i}, X) + \sum (Z_i + F_i^*/X_q^{d_i})^{\mu_i} H_{\mu_i}(Y, Z, X), \end{aligned}$$

where the sum is taken over terms having one factor $(Z_i + F_i^*/X_q^{d_i})$ to some power $\mu_i > 0$, and some factor H_{μ_i} in $\mathbf{Z}[Y, Z, X]$. From the way (y, z, x) was constructed, and the fact that $G(y, z, x) = 0$, we see that the first term vanishes, and hence

$$G(Y, Z, X) = \sum (Z_i + F_i^*/X_q^{d_i})^{\mu_i} H_{\mu_i}(Y, Z, X).$$

Clearing denominators of X_q, for some integer s we get

$$X_q^s G(Y, Z, X) \equiv 0 \bmod (F_i, \ldots, F_r),$$

or in other words, $(*)_q$ is satisfied. This concludes the proof of the theorem.

Remark. Of course the same statement and proof as in Theorem 3.4 holds with $\mathbf{Z}$ replaced by a field k. In that case, we denote by $\mathfrak{a}_k$ the ideal in $k[W, X]$ generated by the generic forms, and similarly by $\mathfrak{p}_k$ the associated prime

ideal. Then

$$\mathfrak{a}_{k,1} = \mathfrak{a}_k \cap k[W] \quad \text{and} \quad \mathfrak{p}_{k,1} = \mathfrak{p}_k \cap k[W].$$

The ideal $\mathfrak{p}$ in Theorem 3.4 will be called the **prime associated with the ideal of generic forms**. The intersection $\mathfrak{p}_1 = \mathfrak{p} \cap \mathbf{Z}[W]$ will be called the prime **elimination ideal** of these forms. If $\mathfrak{A}$ denotes as before the zeros of $\mathfrak{p}$ (or of $\mathfrak{a}$), and $\mathfrak{A}_1$ is its projection on the first factor, then $\mathfrak{p}_1$ is the prime associated with $\mathfrak{A}_1$. The same terminology will be used if instead of $\mathbf{Z}$ we work over a field k. (*Note*: homogeneous elements of $\mathfrak{p}_1$ have been called **inertia forms** in the classical literature, following Hurwitz. I am avoiding this terminology because the word "inertia" is now used in a standard way for inertia groups as in Chapter VII, §2.) The variety of zeros of $\mathfrak{p}_1$ will be called the **resultant variety**. It is determined by the given degrees $d_1, \ldots, d_n$, so we could denote it by $\mathfrak{A}_1(d_1, \ldots, d_n)$.

Exercise. Show that if $\mathfrak{p}$ is the prime associated with the ideal of generic forms, then $\mathfrak{p} \cap \mathbf{Z} = (0)$ is the zero ideal.

Theorem 3.5. *Assume $r = n$, so we deal with n forms in n variables. Then $\mathfrak{p}_1$ is principal, generated by a single polynomial, so $\mathfrak{A}_1$ is what one calls a hypersurface. If (w) is a generic point of $\mathfrak{A}_1$ over a field k, then the transcendence degree of $k(w)$ over k is $m - 1$.*

Proof. We prove the second statement first, and use the same notation as in the proof of Theorem 3.4. Let $u_j = x_j/x_n$. Then $u_n = 1$ and (y), $(u_1, \ldots, u_{n-1})$ are algebraically independent. By (3), we have $z_i = -F_i^*(y, u)$, so

$$k(w) = k(y, z) \subset k(y, u),$$

and so the transcendence degree of $k(w)$ over k is $\leqq m - 1$. We claim that this transcendence degree is $m - 1$. It will suffice to prove that $u_1, \ldots, u_{n-1}$ are algebraic over $k(w) = k(y, z)$. Suppose this is not the case. Then there exists a place φ of $k(w, u)$, which is the identity on $k(w)$ and maps some u_j on ∞. Select an index q such that $\varphi(u_i/u_q)$ is finite for all $i = 1, \ldots, n - 1$. Let $v_i = u_i/u_q$ and $v_i' = \varphi(u_i/u_q)$. Denote by Y_{iq} the coefficient of $X_q^{d_i}$ in F_i and let Y^* denote the variables (Y) from which $Y_{1q}, \ldots, Y_{nq}$ are deleted. By (3) we have for $i = 1, \ldots, n$:

$$\begin{aligned} 0 &= y_{iq}u_q^{d_i} + z_i + F_i^{**}(y^*, u) \\ &= y_{iq} + z_i/u_q^{d_i} + F_i^{**}(y^*, u/u_q). \end{aligned}$$

Applying the place yields

$$0 = y_{iq} + F_i^{**}(y^*, v').$$

In particular, $y_{iq} \in k(y^*, v')$ for each $i = 1, \ldots, n$. But the transcendence degree of $k(v')$ over k is at most $n - 1$, while the elements $(y_{1q}, \ldots, y_{nq}, y^*)$ are algebraically independent over k, which gives a contradiction proving the theorem.

Remark. There is a result (I learned it from [Jo 80]) which is more precise than Theorem 3.5. Indeed, let $\mathfrak{A}$ as in Theorem 3.5 be the variety of zeros of $\mathfrak{p}$, and $\mathfrak{A}_1$ its projection. Then this projection is birational in the following sense. Using the notation of the proof of Theorem 3.5, the result is not only that $k(w)$ has transcendence degree $m - 1$ over k, but actually we have

$$\mathbf{Q}(y, z) = \mathbf{Q}(w) = \mathbf{Q}(y, u).$$

Proof. Let $\mathfrak{p}_1 = (R)$, so R is the resultant, generating the principal ideal $\mathfrak{p}_1$. We shall need the following lemma.

Lemma 3.6. *There is a positive integer s with the following properties. Fix an index i with $1 \leqq i \leqq n - 1$. For each pair of n-tuples of integers $\geqq 0$*

$$(\alpha) = (\alpha_1, \ldots, \alpha_n) \quad \text{and} \quad (\beta) = (\beta_1, \ldots, \beta_n)$$

with $|\alpha| = |\beta| = d_i$, we have

$$X_n^s\left(M_{(\alpha)}(X)\frac{\partial R}{\partial W_{i,(\beta)}} - M_{(\beta)}(X)\frac{\partial R}{\partial W_{i,(\alpha)}}\right) \equiv 0 \bmod (F_1, \ldots, F_n).$$

To see this, we use the fact from Theorem 3.4 that for some s,

$$X_n^s R(W) = Q_1 F_1 + \cdots + Q_n F_n \text{ with } Q_j \in \mathbf{Z}[W, X].$$

Differentiating with respect to $W_{i,(\beta)}$ we get

$$X_n^s \frac{\partial R}{\partial W_{i,(\beta)}} \equiv Q_i M_{(\beta)}(X) \bmod (F_1, \ldots, F_n),$$

and similarly

$$X_n^s \frac{\partial R}{\partial W_{i,(\alpha)}} \equiv Q_i M_{(\alpha)}(X) \bmod (F_1, \ldots, F_n).$$

We multiply the first congruence by $M_{(\alpha)}(X)$ and the second by $M_{(\beta)}(X)$, and we subtract to get our lemma.

From the above we conclude that

$$M_{(\alpha)}(X)\,\frac{\partial R}{\partial W_{i,(\beta)}} - M_{(\beta)}(X)\,\frac{\partial R}{\partial W_{i,(\alpha)}}$$

vanishes on $\mathfrak{A}$, i.e. on the point (w, u), after we put $X_n = 1$. Then we select

$$M_{(\alpha)}(X) = X_i^{d_i} \quad \text{and} \quad M_{(\beta)}(X) = X_i^{d_i - 1}X_n \text{ for } i = 1, \ldots, n - 1,$$

and we see that we have the rational expression

$$u_i = \left.\frac{\partial R/\partial W_{i,(\beta)}}{\partial R/\partial W_{i,(\alpha)}}\right|_{(W)=(w)}, \text{ for } i = 1, \ldots, n - 1,$$

thus showing that $\mathbf{Q}(u) \subset \mathbf{Q}(w)$, as asserted.

We note that the argument also works over the prime field of characteristic p. The only additional remark to be made is that there is some partial derivative $\partial R/\partial W_{i,(\alpha)}$ which does not vanish on (w). This is a minor technical matter, which we leave to the reader.

The above argument is taken from [Jo 80], Proposition 3.3.1. Jouanolou links old-time results as in Macaulay [Ma 16] with more recent techniques of commutative algebra, including the Koszul complex (which will be discussed in Chapter XXI). See also his monographs [Jo 90], [Jo 91].

Still following van der Waerden, we shall now give a fairly explicit determination of the polynomial generating the ideal in Theorem 3.5. We deal with the generic forms $F_i(W, X)$ $(i = 1, \dots, n)$. According to Theorem 3.5, the ideal $\mathfrak{p}_1$ is generated by a single element. Because the units in $\mathbf{Z}[W]$ consist only of ± 1, it follows that this element is well defined up to a sign. Let

$$R(W) = R(F_1, \dots, F_n)$$

be one choice of this element. Later we shall see how to pick in a canonical way one of these two possible choices. We shall prove various properties of this element, which will be called the **resultant** of $F_1, \dots, F_n$.

For each $i = 1, \dots, n$ we let D_i be the product of the degrees with d_i omitted; that is,

$$D_i = d_1 \cdots \hat{d_i} \cdots d_n.$$

We let d be the positive integer such that $d - 1 = \sum (d_i - 1)$.

Lemma 3.7. *Given one of the indices, say n, there is an element $R_n(W)$ lying in $\mathfrak{p}_1$, satisfying the following properties.*

(a) *For each i, $R_n(W)X_i^d \equiv 0 \bmod (F_1, \dots, F_n)$ in $\mathbf{Z}[W, X]$.*
(b) *For each i, $R_n(W)$ is homogeneous in the set of variables $(W_{i,(\nu)})$, and is of degree D_n in $(W_{n,(\nu)})$,* i.e. *in the coefficient of F_n.*
(c) *As a polynomial in $\mathbf{Z}[W]$, $R_n(W)$ has content* 1, i.e. *is primitive.*

Proof. The polynomial $R_n(W)$ will actually be explicitly constructed. Let $M_\sigma(X)$ denote the monomials of degree $|\sigma| = d$. We partition the indexing set $S = \{\sigma\}$ into disjoint subsets as follows.

Let $S_1 = \{\sigma_1\}$ be the set of indices such that $M_{\sigma_1}(X)$ is divisible by $X_1^{d_1}$.

Let $S_2 = \{\sigma_2\}$ be the set of indices such that $M_{\sigma_2}(X)$ is divisible by $X_2^{d_2}$ but not by $X_1^{d_1}$.

. . .

Let $S_n = \{\sigma_n\}$ be the set of indices such that $M_{\sigma_n}(X)$ is divisible by $X_n^{d_n}$ but not by $X_1^{d_1}, \dots, X_{n-1}^{d_{n-1}}$.

Then S is the disjoint union of $S_1, \dots, S_n$. Write each monomial as follows:

$$\begin{aligned} M_{\sigma_1}(X) &= H_{\sigma_1}(X)X_1^{d_1} \quad \text{so} \quad \deg H_{\sigma_1} = d - d_1 \\ &\vdots \qquad \vdots \\ M_{\sigma_1}(X) &= H_{\sigma_n}(X)X_n^{d_n} \quad \text{so} \quad \deg H_{\sigma_n} = d - d_n. \end{aligned}$$

Then the number of polynomials

$$H_{\sigma_1}F_1, \dots, H_{\sigma_n}F_n \text{ (with } \sigma_1 \in S_1, \dots, \sigma_n \in S_n)$$

is precisely equal to the number of monomials of degree d. We let R_n be the determinant of the coefficients of these polynomials, viewed as forms in (X) with coefficients in $\mathbf{Z}[W]$. Then $R_n = R_n(W) \in \mathbf{Z}[W]$. We claim that $R_n(W)$ satisfies the properties of the lemma.

First we note that if $\sigma_n \in S_n$, then $H_{\sigma_n}(X)$ is divisible by a power of X_i at most $d_i - 1$, for $i = 1, \dots, n - 1$. On the other hand, the degree of $H_{\sigma_n}(X)$ in X_n is determined by the condition that the total degree is $d - d_n$. Hence S_n has exactly D_n elements. It follows at once that $R_n(W)$ is homogeneous of degree D_n in the coefficients of F_n, i.e. in $(W_{n,(\nu)})$. From the construction it also follows that R_n is homogeneous in each set of variables $(W_{i,(\nu)})$ for each $i = 1, \dots, n - 1$.

If we specialize the forms F_i $(i = 1, \dots, n)$ to $X_i^{d_i}$, then R_n specializes to 1, and hence $R_n \neq 0$ and R_n is primitive. For each σ_i we can write

$$H_{\sigma_i}F_i = \sum_{\sigma \in S} C_{\sigma,\sigma_i}(W)M_\sigma(X),$$

where $M_\sigma(X)$ $(\sigma \in S)$ ranges over all monomials of degree d in (X), and $C_{\sigma,\sigma_i}(W)$ is one of the variables (W). Then by definition

$$R_n(W) = \det(C_{\sigma,\sigma_1}(W)_{(\sigma_1 \in S_1)}, \dots, C_{\sigma,\sigma_n}(W)_{(\sigma_n \in S_n)}) = \det(C).$$

where $\sigma_1 \in S_1, \dots, \sigma_n \in S_n$ indexes the columns, and σ indexes the rows. Let $B = \tilde{C}$ be the matrix with components in $\mathbf{Z}[W, X]$ such that

$$BC = \det(C)I = R_nI.$$

(See Chapter XIII, Corollary 4.17.) Then for each σ, we have

$$R_n(W)M_\sigma(X) = \sum_i \sum_{\sigma_i \in S_i} B_{i,\sigma_i}F_i.$$

Given i, we take for σ the index such that $M_\sigma(X) = X_i^d$ in order to obtain the first relation in Lemma 3.7. By Theorem 3.4, we conclude that $R_n(W) \in \mathfrak{p}_1$. This concludes the proof of the lemma.

Of course, we picked an index n to fix ideas. For each i one has a polynomial R_i satisfying the analogous properties, and in particular homogeneous of degree D_i in the variables $(W_{i,(\nu)})$ which are the coefficients of the form F_i.

Theorem 3.8. *Let R be the resultant of the n generic forms F_i over $\mathbf{Z}$, in n variables. Then R satisfies the following properties.*

(a) *R is the greatest common divisor in $\mathbf{Z}[W]$ of the polynomials $R_1, \ldots, R_n$.*
(b) *R is homogeneous of degree D_i in the coefficients of F_i.*
(c) *Let $F_i = \ldots + W_{i,(d_i)}X_i^{d_i}$, so $W_{i,(d_i)}$ is the coefficient of $X_i^{d_i}$. Then R contains the monomial*

$$\pm \prod_{i=1}^{n} W_{i,(d_i)}^{D_i}.$$

Proof. The idea will be to specialize the forms $F_1, \ldots, F_n$ to products of generic linear forms, where we can tell what is going on. For that we need a lemma of a more general property eventually to be proved. We shall use the following notation. If $f_1, \ldots, f_n$ are forms with coefficients (w), then we write

$$R(f_1, \ldots, f_n) = R(w).$$

Lemma 3.9. *Let G, H be generic independent forms with* $\deg(GH) = d_1$. *Then $R(GH, F_2, \ldots, F_n)$ is divisible by $R(G, F_2, \ldots, F_n)R(H, F_2, \ldots, F_n)$.*

Proof. By Theorem 3.5, there is an expression

$$X_n^s R(F_1, \ldots, F_n) = Q_1 F_1 + \cdots + Q_n F_n \text{ with } Q_i \in \mathbf{Z}[W, X].$$

Let $W_G, W_H, W_{F_2}, \ldots, W_{F_n}$ be the coefficients of $G, H, F_2, \ldots, F_n$ respectively, and let (w) be the coefficients of $GH, F_2, \ldots, F_n$. Then

$$R(w) = R(GH, F_2, \ldots, F_n),$$

and we obtain

$$X_n^s R(w) = Q_1(w, X)GH + Q_2(w, X)F_2 + Q_n(w, X)F_n.$$

Hence $R(GH, F_2, \ldots, F_n)$ belongs to the elimination ideal of $G, F_2, \ldots, F_n$ in the ring $\mathbf{Z}[W_G, W_H, W_{F_2}, \ldots, W_{F_n}]$, and similarly with H instead of G. Since W_H is a family of independent variables over $\mathbf{Z}[W_G, W_{F_2}, \ldots, W_{F_n}]$, it follows that $R(G, F_2, \ldots, F_n)$ divides $R(GH, F_2, \ldots, F_n)$ in that ring, and similarly for $R(H, F_2, \ldots, F_n)$. But (W_G) and (W_H) are independent sets of variables, and so $R(G, F_2, \ldots, F_n)$, $R(H, F_2, \ldots, F_n)$ are distinct prime elements in that ring, so their product divides $R(GH, F_2, \ldots, F_n)$ as stated, thus proving the lemma.

Lemma 3.9 applies to any specialized family of polynomials $g, h, f_1, \ldots, f_n$ with coefficients in a field k. Observe that for a system of n linear forms in n variables, the resultant is simply the determinant of the coefficients. Thus if $L_1, \ldots, L_n$ are generically independent linear forms in the variables $X_1, \ldots, X_n$, then their resultant $R(L_1, \ldots, L_n)$ is homogeneous of degree 1 in the coefficients of L_i for each i. We apply Lemma 3.9 to the case of forms $f_1, \ldots, f_{n-1}$, which are products of generically independent linear forms. By Lemma 3.9 we conclude that for this specialized family of form, their resultant has degree at least D_n in

the coefficients of F_n, so for the generic forms $F_1, \ldots, F_n$ their resultant has degree at least D_n in the coefficients of F_n. Similarly $R(F_1, \ldots, F_n)$ has degree at least D_i in the coefficients of F_i for each i. But R divides the n elements $R_1(W), \ldots, R_n(W)$ constructed in Lemma 3.7. Therefore we conclude that R has degree exactly D_i in the coefficients of F_i. By Theorem 3.5, we know that R divides each R_i. Let G be the greatest common divisor of $R_1, \ldots, R_n$ in $\mathbf{Z}[W]$. Then R divides G and has the same degree in each set of variables $(W_{i,(\nu)})$ for $i = 1, \ldots, n$. Hence there exists $c \in \mathbf{Z}$ such that $G = cR$. We must have $c = \pm 1$, because, say, R_n is primitive in $\mathbf{Z}[W]$. This proves (a) and (b) of the theorem.

As to the third part, we specialize the forms to $f_i = X_i^{d_i}$, $i = 1, \ldots, n$. Then R_n specializes to 1, and since R divides R_n it follows that R itself specializes to ± 1. Since all coefficients of the forms specialize to 0 except those which we denoted by $W_{i,(d_i)}$, it follows that $R(W)$ contains the monomial which is the product of these variables to the power D_i, up to the sign ± 1. This proves (c), and concludes the proof of Theorem 3.8.

We can now normalize the resultant by choosing the sign such that R contains the monomial

$$M = \prod_{i=1}^{n} W_{i,(d_i)}^{D_i},$$

with coefficient $+1$. This condition determines R uniquely, and we then denote R also by

$$R = \operatorname{Res}(F_1, \ldots, F_n).$$

Given forms $f_1, \ldots, f_n$ with coefficients (w) in a field K (actually any commutative ring), we can then define their **resultant**

$$\operatorname{Res}(f_1, \ldots, f_n) = R(w)$$

with the normalized polynomial R. With this normalization, we then have a stronger result than Lemma 3.9.

Theorem 3.10. *Let $f_1 = gh$ be a product of forms such that* $\deg(gh) = d_1$. *Let $f_2, \ldots, f_n$ be arbitrary forms of degrees $d_2, \ldots, d_n$. Then*

$$\operatorname{Res}(gh, f_2, \ldots, f_n) = \operatorname{Res}(g, f_2, \ldots, f_n)\operatorname{Res}(h, f_2, \ldots, f_n).$$

Proof. From the fact that the degrees have to add in a product of polynomials, together with Theorem 3.8(a) and (b), we now see in Lemma 3.9 that we must have the precise equality in what was only a divisibility before we knew the precise degree of R in each set of variables.

Theorem 3.10 is very useful in proving further properties of the determinant, because it allows a reduction to simple cases under factorization of polynomials.

For instance one has:

Theorem 3.11. *Let $F_1, \ldots, F_n$ be the generic forms in n variables, and let $\overline{F}_1, \ldots, \overline{F}_n$ be the forms obtained by substituting $X_n = 0$, so that $\overline{F}_1, \ldots, \overline{F}_{n-1}$ are the generic forms in $n - 1$ variables. Let $n \geqq 2$. Then*

$$\operatorname{Res}(F_1, \ldots, F_{n-1}, X_n^{d_n}) = \operatorname{Res}(\overline{F}_1, \ldots, \overline{F}_{n-1})^{d_n}.$$

Proof. By Theorem 3.10 it suffices to prove the assertion when $d_n = 1$. By Theorem 3.4, for each $i = 1, \ldots, n - 1$ we have an expression

$$(*) \qquad X_i^s \operatorname{Res}(F_1, \ldots, F_{n-1}, X_n) = Q_1 F_1 + \cdots + Q_{n-1}F_{n-1} + Q_n X_n$$

with $Q_j \in \mathbf{Z}[W, X]$ (depending on the choice of i). The left-hand side can be written as a polynomial in the coefficients of $F_1, \ldots, F_{n-1}$ with the notation

$$X_i^s R(W_{F_1}, \ldots, W_{F_{n-1}}, 1_{X_n}) = X_i^s P(W_{F_1}, \ldots, W_{F_{n-1}}) = X_i^s P(W^{(n-1)}), \text{ say;}$$

thus in the generic linear form in $X_1, \ldots, X_n$ we have specialized all the coefficients to 0 except the coefficient of X_n, which we have specialized to 1. Substitute $X_n = 0$ in the right side of (*). By Theorem 3.4, we conclude that $P(W^{(n-1)})$ lies in the resultant ideal of $\overline{F}_1, \ldots, \overline{F}_{n-1}$, and therefore $\operatorname{Res}(\overline{F}_1, \ldots, \overline{F}_{n-1})$ divides $P(W^{(n-1)})$. By Theorem 3.8 we know that $P(W^{(n-1)})$ has the same homogeneity degree in $W_{\overline{F}_i}$ $(i = 1, \ldots, n - 1)$ as $\operatorname{Res}(\overline{F}_1, \ldots, \overline{F}_{n-1})$. Hence there is $c \in \mathbf{Z}$ such that

$$c\operatorname{Res}(\overline{F}_1, \ldots, \overline{F}_{n-1}) = \operatorname{Res}(F_1, \ldots, F_{n-1}, X_n).$$

One finds $c = 1$ by specializing $\overline{F}_1, \ldots, \overline{F}_{n-1}$ to $X_1^{d_1}, \ldots, X_{n-1}^{d_{n-1}}$ respectively, thus concluding the proof.

The next basic lemma is stated for the generic case, for instance in Macaulay [Ma 16], and is taken up again in [Jo 90], Lemma 5.6.

Lemma 3.12. *Let A be a commutative ring. Let $f_1, \ldots, f_n, g_1, \ldots, g_n$ be homogeneous polynomials in $A[X_1, \ldots, X_n]$. Assume that*

$$(g_1, \ldots, g_n) \subset (f_1, \ldots, f_n)$$

as ideals in $A[X]$. Then

$$\operatorname{Res}(f_1, \ldots, f_n) \text{ divides } \operatorname{Res}(g_1, \ldots, g_n) \text{ in } A.$$

Proof. Express each $g_i = \sum h_{ij} f_j$ with h_{ij} homogeneous in $A[X]$. By specialization, we may then assume that $g_i = \sum H_{ij} F_j$ where H_{ij} and F_j have algebraically independent coefficients over $\mathbf{Z}$. By Theorem 3.4, for each i we have a relation

$$X_i^s \operatorname{Res}(g_1, \ldots, g_n) = Q_1 g_1 + \cdots + Q_n g_n \text{ with some } Q_i \in \mathbf{Z}[W_H, W_F],$$

where W_H, W_F denote the independent variable coefficients of the polynomials H_{ij} and F_j respectively. In particular,

(*) $$X_i^s \operatorname{Res}(g_1, \ldots, g_n) \equiv 0 \bmod (F_1, \ldots, F_n)\mathbf{Z}[W_H, W_F, X].$$

Note that $\operatorname{Res}(g_1, \ldots, g_n) = P(W_H, W_F) \in \mathbf{Z}[W_H, W_F]$ is a polynomial with integer coefficients. If (w_F) is a generic point of the resultant variety $\mathfrak{A}_1$ over $\mathbf{Z}$, then $P(W_H, w_F) = 0$ by (*). Hence $\operatorname{Res}(F_1, \ldots, F_n)$ divides $P(W_H, W_F)$, thus proving the lemma.

Theorem 3.13. *Let A be a commutative ring and let $d_1, \ldots, d_n$ be integers $\geqq 1$ as usual. Let f_i be homogeneous of degree d_i in $A[X] = A[X_1, \ldots, X_n]$. Let d be an integer $\geqq 1$, and let $g_i, \ldots, g_n$ be homogeneous of degree d in $A[X]$. Then*

$$f_i \circ g = f_i(g_1, \ldots, g_n)$$

is homogeneous of degree dd_i, and

$$\operatorname{Res}(f_1 \circ g, \ldots, f_n \circ g) = \operatorname{Res}(g_1, \ldots, g_n)^{d_1 \cdots d_n} \operatorname{Res}(f_1, \ldots, f_n)^{d^{n-1}} \text{ in } A.$$

Proof. We start with the standard relation of Theorem 3.4:

(*) $$X_i^s \operatorname{Res}(F_1, \ldots, F_n) \equiv 0 \bmod (F_1, \ldots, F_n)\mathbf{Z}[W_F, X].$$

We let $G_1, \ldots, G_n$ be independent generic polynomials of degree d, and let W_G denote their independent variable coefficients. Substituting G_i for X_i in (*), we find

$$G_i^s \operatorname{Res}(F_1, \ldots, F_n) \equiv 0 \bmod (F_1 \circ G, \ldots, F_n \circ G)\mathbf{Z}[W_F, W_G, X].$$

Abbreviate $\operatorname{Res}(F_1, \ldots, F_n)$ by $R(F)$, and let $g_i = G_i^s R(F)$. By Lemma 3.12, it follows that

$$\operatorname{Res}(f_1 \circ G, \ldots, F_n \circ G) \text{ divides } \operatorname{Res}(G_1^s R(F), \ldots, G_n^s R(F)) \text{ in } \mathbf{Z}[W_F, W_G].$$

By Theorem 3.10 and the homogeneity of Theorem 3.8(b) we find that

$$\operatorname{Res}(G_1^s R(F), \ldots, G_n^s R(F)) = \operatorname{Res}(G_1, \ldots, G_n)^M \operatorname{Res}(F_1, \ldots, F_n)^N$$

with integers $M, N \geqq 0$. Since $\operatorname{Res}(G_1, \ldots, G_n)$ and $\operatorname{Res}(F_1, \ldots, F_n)$ are distinct prime elements in $\mathbf{Z}[W_G, W_F]$ (distinct because they involve independent variables), it follows that

(**) $$\operatorname{Res}(F_1 \circ G, \ldots, F_n \circ G) = \varepsilon \operatorname{Res}(G_1, \ldots, G_n)^a \operatorname{Res}(F_1, \ldots, F_n)^b$$

with integers $a, b \geqq 0$ and $\varepsilon = 1$ or -1. Finally, we specialize F_i to $W_i X_i^{d_i}$ and we specialize G_i to $U_i X_i^d$, with independent variables $(W_1, \ldots, W_n, U_1, \ldots, U_n)$.

Substituting in (**), we obtain

$$\operatorname{Res}(W_1 U_1^{d_1} X_1^{dd_1}, \ldots, W_n U_n^{d_n} X_n^{dd_n})$$
$$= \varepsilon \operatorname{Res}(U_1 X_1^d, \ldots, U_n X_n^d)^a \operatorname{Res}(W_1 X_1^{d_1}, \ldots, W_n X_n^{d_n})^b.$$

By the homogeneity of Theorem 3.8(b) we get

$$\prod_i (W_i U^{d_i})^{d_1 \cdots \hat{d}_i \cdots d_n d^{n-1}} = \varepsilon \prod_i U_i^{d^{n-1} a} \prod_i W_i^{d_1 \cdots \hat{d}_i \cdots d_n b}.$$

From this we get at once $\varepsilon = 1$ and a, b are what they are stated to be in the theorem.

Corollary 3.14. *Let $C = (c_{ij})$ be a square matrix with coefficients in A. Let $f_i(X) = F_i(CX)$ (where CX is multiplication of matrices, viewing X as a column vector). Then*

$$\operatorname{Res}(f_1, \ldots, f_n) = \det(C)^{d_1 \cdots d_n} \operatorname{Res}(F_1, \ldots, F_n).$$

Proof. This is the case when $d = 1$ and g_i is a linear form for each i.

Theorem 3.15. *Let $f_1, \ldots, f_n$ be homogeneous in $A[X]$, and suppose $d_n \geqq d_i$ for all i. Let h_i be homogeneous of degree $d_n - d_i$ in $A[X]$. Then*

$$\operatorname{Res}(f_1, \ldots, f_{n-1}, f_n + \sum_{j=1}^{n-1} h_j f_j) = \operatorname{Res}(f_1, \ldots, f_n) \text{ in } A.$$

Proof. We may assume $f_i = F_i$ are the generic forms, H_i are forms generic independent from $F_1, \ldots, F_n$, and $A = \mathbf{Z}[W_F, W_H]$, where (W_F) and (W_H) are the coefficients of the respective polynomials. We note that the ideals $(F_1, \ldots, F_n)$ and $(F_1, \ldots, F_n + \sum_{j \neq n} H_j F_j)$ are equal. From Lemma 3.12 we conclude that the two resultants in the statement of the theorem differ by a factor of 1 or -1. We may now specialize H_{ij} to 0 to determine that the factor is $+1$, thus concluding the proof.

Theorem 3.16. *Let π be a permutation of $\{1, \ldots, n\}$, and let $\varepsilon(\pi)$ be its sign. Then*

$$\operatorname{Res}(F_{\pi(1)}, \ldots, F_{\pi(n)}) = \varepsilon(\pi)^{d_1 \cdots d_n} \operatorname{Res}(F_1, \ldots, F_n).$$

Proof. Again using Lemma 3.12 with the ideals $(F_1, \ldots, F_n)$ and $(F_{\pi(1)}, \ldots, F_{\pi(n)})$, which are equal, we conclude the desired equality up to a factor ± 1, in $\mathbf{Z}[W_F]$. We determine this sign by specializing F_i to $X_i^{d_i}$, and using the multiplicativity of Theorem 3.10. We are then reduced to the case when $F_i = X_i$, so a linear form; and we can apply Corollary 3.14 to conclude the proof.

The next theorem was an exercise in van der Waerden's *Moderne Algebra*.

Theorem 3.17. *Let $L_1, \ldots, L_{n-1}, F$ be generic forms in n variables, such that $L_1, \ldots, L_{n-1}$ are of degree 1, and F has degree $d = d_n$. Let*

$$\Delta_j (j = 1, \ldots, n)$$

be $(-1)^{n-j}$ times the j-th minor determinant of the coefficient matrix of the forms $(L_1, \ldots, L_{n-1})$. Then

$$\operatorname{Res}(L_1, \ldots, L_{n-1}, F) = F(\Delta_1, \ldots, \Delta_n).$$

Proof. We first claim that for all $j = 1, \ldots, n$ we have the congruence

$$(*) \qquad X_n \Delta_j - X_j \Delta_n \equiv 0 \bmod (L_1, \ldots, L_{n-1})\mathbf{Z}[W, X],$$

where as usual, (W) are the coefficients of the forms $L_1, \ldots, L_{n-1}, F$. To see this, we consider the system of linear equations

$$\begin{gathered} W_{11}X_1 + \cdots + W_{1,n-1}X_{n-1} = L_1(W, X) - W_{1,n}X_n \\ \cdots\cdots \\ W_{n-1,1}X_1 + \cdots + W_{n-1,n-1}X_{n-1} = L_{n-1}(W, X) - W_{n-1,n}X_n . \end{gathered}$$

If $C = (C^1, \ldots, C^{n-1})$ is a square matrix with columns C^j, then a solution of a system of linear equations $CX = C^n$ satisfies Cramer's rule

$$X_j \det(C^1, \ldots, C^{n-1}) = \det(C^1, \ldots, C^n, \ldots, C^{n-1}).$$

Using the fact that the determinant is linear in each column, (*) falls out.
Then from the congruence (*) it follows that

$$X_n^d F(\Delta_1, \ldots, \Delta_n) \equiv \Delta_n^d F(X_1, \ldots, X_n) \bmod (L_1, \ldots, L_{n-1})\mathbf{Z}[W, X],$$

whence

$$X_n^d F(\Delta_1, \ldots, \Delta_n) \equiv 0 \bmod (L_1, \ldots, L_{n-1}, F).$$

Hence by Theorem 3.4 and the fact that $\operatorname{Res}(L_1, \ldots, L_{n-1}, F) = R(W)$ generates the elimination ideal, it follows that there exists $c \in \mathbf{Z}[W]$ such that

$$F(\Delta_1, \ldots, \Delta_n) = c\operatorname{Res}(L_1, \ldots, L_{n-1}, F).$$

Since the left side is homogeneous of degree 1 in the coefficients W_F and homogeneous of degree d in the coefficients W_{L_i} for each $i = 1, \ldots, n-1$, it follows from Theorem 3.8 that $c \in \mathbf{Z}$. Specializing L_i to X_i and F to X_n^d makes Δ_j specialize to 0 if $j \neq n$ and Δ_n specializes to 1. Hence the left side specializes to 1, and so does the right side, whence $c = 1$. This concludes the proof.

Bibliography

[Jo 80] J. P. JOUANOLOU, Idéaux résultants, *Advances in Mathematics* **37** No. 3 (1980), pp. 212–238

[Jo 90] J. P. JOUANOLOU, Le formalisme du résultant, *Advances in Mathematics* **90** No. 2 (1991) pp. 117–263

[Jo 91] J. P. JOUANOLOU, *Aspects invariants de l'élimination*, Départment de Mathématiques, Université Louis Pasteur, Strasbourg, France (1991)

[Ma 16] F. MACAULAY, *The algebraic theory of modular systems*, Cambridge University Press, 1916

§4. RESULTANT SYSTEMS

The projection argument used to prove Theorem 3.4 has the advantage of constructing a generic point in a very explicit way. On the other hand, no explicit, or even effective, formula was given to construct a system of forms defining $\mathfrak{A}_1$. We shall now reformulate a version of Theorem 3.4 over $\mathbf{Z}$ and we shall prove it using a completely different technique which constructs effectively a system of generators for an ideal of definition of the arithmetic variety $\mathfrak{A}_1$ in Theorem 3.2.

Theorem 4.1. *Given degrees $d_1, \ldots, d_r \geqq 1$, and positive integers m, n. Let $(W) = (W_{i,(\nu)})$ be the variables as in* §3, (2) *viewed as algebraically independent elements over the integers $\mathbf{Z}$. There exists an effectively determinable finite number of polynomials $R_\rho(W) \in \mathbf{Z}[W]$ having the following property. Let (f) be as in* (1), *a system of forms of the given degrees with coefficients (w) in some field k. Then (f) has a non-trivial common zero if and only if $R_\rho(w) = 0$ for all ρ.*

A finite family $\{R_\rho\}$ having the property stated in Theorem 4.1 will be called a **resultant system** for the given degrees. According to van der Waerden (*Moderne Algebra*, first and second edition, §80), the following technique of proof using resultants goes back to Kronecker elimination, and to a paper of Kapferer (Über Resultanten und Resultantensysteme, *Sitzungsber. Bayer. Akad. München* 1929, pp. 179–200). The family of polynomials $\{R_\rho(W)\}$ is called a **resultant system**, because of the way they are constructed. They form a set of generators for an ideal $\mathfrak{b}_1$ such that the arithmetic variety $\mathfrak{A}_1$ is the set of zeros of $\mathfrak{b}_1$. I don't know how close the system constructed below is to being a set of generators for the prime ideal $\mathfrak{p}_1$ in $\mathbf{Z}[W]$ associated with $\mathfrak{A}_1$. Actually we shall not need the whole theory of Chapter IV, §10; we need only one of the characterizing properties of resultants.

Let p, q be positive integers. Let

$$f_v = v_0X_1^p + v_1X_1^{p-1}X_2 + \cdots + v_pX_2^p$$
$$g_w = w_0X_1^q + w_1X_1^{q-1}X_2 + \cdots + w_qX_2^q$$

be two generic homogeneous polynomials in $\mathbf{Z}[v, w, X_1, X_2] = \mathbf{Z}[v, w][X]$. In Chapter IV, §10 we defined their resultant $\mathrm{Res}(f_v, g_w)$ in case $X_2 = 1$, but we find it now more appropriate to work with homogeneous polynomials. For our purposes here, we need only the fact that the resultant $R(v, w)$ is characterized by the following property. If we have a specialization (a, b) of (v, w) in a field K, and if f_a, f_b have a factorization

$$f_a = a_0 \prod_{i=1}^{p} (X_1 - \alpha_i X_2)$$

$$g_b = b_0 \prod_{j=1}^{q} (X_1 - \beta_j X_2)$$

then we have the symmetric expressions in terms of the roots:

$$\begin{aligned} R(a, b) = \mathrm{Res}(f_a, f_b) &= a_0^q b_0^p \prod_{i,j} (\alpha_i - \beta_j) \\ &= a_0^q \prod_i g_b(\alpha_i, 1) = (-1)^{pq} b_0^p \prod_j f_a(\beta_j, 1). \end{aligned}$$

From the general theory of symmetric polynomials, it is *a priori* clear that $R(v, w)$ lies in $\mathbf{Z}[v, w]$, and Chapter IV, §10 gives an explicit representation

$$\varphi_{v,w} f_v + \psi_{v,w} g_w = X_2^{p+q-1} R(v, w)$$

where $\varphi_{v,w}$ and $\psi_{v,w} \in \mathbf{Z}[v, w, X]$. This representation will not be needed. The next property will provide the basic inductive step for elimination.

Proposition 4.2. *Let f_a, g_b be homogeneous polynomials with coefficients in a field K. Then $R(a, b) = 0$ if and only if the system of equations*

$$f_a(X) = 0, \quad g_b(X) = 0$$

has a non-trivial zero in some extension of K (which can be taken to be finite).

If $a_0 = 0$ then a zero of g_b is also a zero of f_a; and if $b_0 = 0$ then a zero of f_a is also a zero of g_b. If $a_0b_0 \neq 0$ then from the expression of the resultant as a product of the difference of roots $(\alpha_i - \beta_j)$ the proposition follows at once.

We shall now prove Theorem 4.1 by using resultants. We do this by induction on n.

If $n = 1$, the theorem is obvious.

If $n = 2$, $r = 1$, the theorem is again obvious, taking the empty set for (R_ρ).

If $n = 2$, $r = 2$, then the theorem amounts to Proposition 4.2.

Assume now $n = 2$ and $r > 2$, so we have a system of homogeneous equations

$$0 = f_1(X) = f_2(X) = \ldots = f_r(X)$$

with $(X) = (X_1, X_2)$. Let d_i be the degree of f_i and let $d = \max d_i$. We replace the family $\{f_j(X)\}$ by the family of all polynomials

$$f_i(X)X_1^{d-d_i} \quad \text{and} \quad f_i(X)X_2^{d-d_i}, \; i = 1, \ldots, r.$$

These two families have the same sets of non-trivial zeros, so to prove Theorem 4.1 we may assume without loss of generality that all the polynomials $f_1, \ldots, f_r$ have the same degree d.

With $n = 2$, consider the generic system of forms of degree d in (X):

(4) $\quad F_i(W, X) = 0$ with $i = 1, \ldots, r$, in two variables $(X) = (X_1, X_2)$,

where the coefficients of F_i are $W_{i,0}, \ldots, W_{i,d}$ so that

$$(W) = (W_{1,0}, \ldots, W_{1,d}, \ldots, W_{r,0}, \ldots, W_{r,d}).$$

The next proposition is a special case of Theorem 4.1, but gives the first step of an induction showing how to get the analogue of Proposition 4.2 for such a larger system. Let $T_1, \ldots, T_r$ and $U_1, \ldots, U_r$ be independent variables over $\mathbf{Z}[W, X]$. Let $F_1, \ldots, F_r$ be the generic forms of §3, (2). Let

$$f = F_1(W, X)T_1 + \cdots + F_r(W, X)T_r$$
$$g = F_1(W, X)U_1 + \cdots + F_r(W, X)U_r$$

so $f, g \in \mathbf{Z}[W, T, U][X]$. Then f, g are polynomials in (X) with coefficients in $\mathbf{Z}[W, T, U]$. We may form their resultant

$$\operatorname{Res}(f, g) \in \mathbf{Z}[W, T, U].$$

Thus $\operatorname{Res}(f, g)$ is a polynomial in the variables (T, U) with coefficients in $\mathbf{Z}[W]$. We let $(Q_\mu(W))$ be the family of coefficients of this polynomial.

Proposition 4.3. *The system $\{Q_\mu(W)\}$ just constructed satisfies the property of Theorem 4.1,* i.e. *it is a resultant system for r forms of the same degree d.*

Proof. Suppose that there is a non-trivial solution of a special system $F_j(W, X) = 0$ with (w) in some field k. Then (w, T, U) is a common non-trivial zero of f, g, so $\operatorname{Res}(f, g) = 0$ and therefore $Q_\mu(w) = 0$ for all μ. Conversely, suppose that $Q_\mu(w) = 0$ for all μ. Let $f_i(X) = F_i(w, X)$. We want to show that $f_i(X)$ for $i = 1, \ldots, r$ have a common non-trivial zero in some extension of

k. If all f_i are 0 in $k[X_1, X_2]$ then they have a common non-trivial zero. If, say, $f_1 \neq 0$ in $k[X]$, then specializing $T_2, \ldots, T_r$ to 0 and T_1 to 1 in the resultant $\mathrm{Res}(f, g)$, we see that

$$\mathrm{Res}(f_1, f_2U_2 + \cdots + f_rU_r) = 0$$

as a polynomial in $k[U_2, \ldots, U_r]$. After making a finite extension of k if necessary, we may assume that $f_1(X)$ splits into linear factors. Let $\{\alpha_i\}$ be the roots of $f_1(X_1, 1)$. Then some $(\alpha_i, 1)$ must also be a zero of $f_2U_2 + \cdots + f_rU_r$, which implies that $(\alpha_i, 1)$ is a common zero of $f_1, \ldots, f_r$ since $U_2, \ldots, U_r$ are algebraically independent over k. This proves Proposition 4.3.

We are now ready to do the inductive step with $n > 2$. Again, let

$$f_i(X) = F_i(w, X) \text{ for } j = 1, \ldots, r$$

be polynomials with coefficients (w) in some fields k.

Remark 4.4. *There exists a non-trivial zero of the system*

$$f_i = 0 \ (i = 1, \ldots, r)$$

in some extension of k if and only if there exist

$$(x_1, \ldots, x_{n-1}) \neq (0, \ldots, 0) \text{ and } (x_n, t) \neq (0, 0)$$

in some extension of k such that

$$f_i(tx_1, \ldots, tx_{n-1}, x_n) = 0 \text{ for } i = 1, \ldots, r.$$

So we may now construct the system (R_ρ) inductively as follows.

Let T be a new variable, and let $X^{(n-1)} = (X_1, \ldots, X_{n-1})$. Let

$$g_i(W, X^{(n-1)}, S_n, T) = F_i(W, TX_1, \ldots, TX_{n-1}, X_n) \in \mathbf{Z}[W, X^{(n-1)}][X_n, T].$$

Then g_i is homogeneous in the two variables (X_n, T). By the theorem for two variables, there is a system of polynomials (Q_μ) in $\mathbf{Z}[W, X^{(n-1)}]$ having the property: *if $(w, x^{(n-1)})$ is a point in a field K, then*

$$g_i(w, x^{(n-1)}, X_n, T) \textit{ have a non-trivial common zero for } i = 1, \ldots, r.$$

$$\Leftrightarrow Q_\mu(w, x^{(n-1)}) = 0 \textit{ for all } \mu.$$

Viewing each Q_μ as a polynomial in the variables $(X^{(n-1)})$, we decompose each Q_μ as a sum of its homogeneous terms, and we let $(H_\lambda(W, X^{(n-1)}))$ be the family of these polynomials, homogeneous in $(X^{(n-1)})$. From the homogeneity property of the forms F_j in (X), it follows that if t is transcendental over K and $g_i(w, x^{(n-1)}, X_n, T)$ have a non-trivial common zero for $j = 1, \ldots, r$ then $g_i(w, tx^{(n-1)}, X_n, T)$ also have a non-trivial common zero. Therefore

$Q_\mu(w, tx^{(n-1)}) = 0$ for all μ, and so $H_\lambda(w, x^{(n-1)}) = 0$. Therefore we may use the family of polynomials (H_λ) instead of the family (Q_μ), and we obtain the property: *if* $(w, x^{(n-1)})$ *is a point in a field* K, *then*

$$g_i(w, x^{(n-1)}, X_n, T) \text{ have a non-trivial common zero for } i = 1, \ldots, r$$

$$\Leftrightarrow H_\lambda(w, x^{(n-1)}) = 0 \text{ for all } \lambda.$$

By induction on n, there exists a family $(R_\rho(W))$ of polynomials in $\mathbf{Z}[W]$ (actually homogeneous), having the property: *if* (w) *is a point in a field* K, *then*

$$H_\lambda(w, X^{(n-1)}) \text{ have a non-trivial common zero for all } \lambda$$

$$\Leftrightarrow R_\rho(w) = 0 \text{ for all } \rho.$$

In light of Remark 4.4, this concludes the proof of Theorem 4.1 by the resultant method.

§5. SPEC OF A RING

We shall extend the notions of §2 to arbitrary commutative rings.

Let A be a commutative ring. By spec(A) we mean the set of all prime ideals of A. An element of spec(A) is also called a **point** of spec(A).

If $f \in A$, we view the set of prime ideals $\mathfrak{p}$ of spec(A) containing f as the set of **zeros** of f. Indeed, it is the set of $\mathfrak{p}$ such that the image of f in the canonical homomorphism

$$A \to A/\mathfrak{p}$$

is 0. Let $\mathfrak{a}$ be an ideal, and let $\mathfrak{X}(\mathfrak{a})$ (the set of **zeros** of $\mathfrak{a}$) be the set of all primes of A containing $\mathfrak{a}$. Let $\mathfrak{a}, \mathfrak{b}$ be ideals. Then we have:

Proposition 5.1.

(i) $\mathfrak{X}(\mathfrak{a}\mathfrak{b}) = \mathfrak{X}(\mathfrak{a}) \cup \mathfrak{X}(\mathfrak{b})$.

(ii) *If* $\{\mathfrak{a}_i\}$ *is a family of ideals, then* $\mathfrak{X}(\sum \mathfrak{a}_i) = \bigcap \mathfrak{X}(\mathfrak{a}_i)$.

(iii) *We have* $\mathfrak{X}(\mathfrak{a}) \subset \mathfrak{X}(\mathfrak{b})$ *if and only if* $\operatorname{rad}(\mathfrak{a}) \supset \operatorname{rad}(\mathfrak{b})$, *where* $\operatorname{rad}(\mathfrak{a})$, *the radical of* $\mathfrak{a}$, *is the set of all elements* $x \in A$ *such that* $x^n \in \mathfrak{a}$ *for some positive integer* n.

Proof. Exercise. See Corollary 2.3 of Chapter X.

A subset C of spec(A) is said to be **closed** if there exists an ideal $\mathfrak{a}$ of A such that C consists of those prime ideals $\mathfrak{p}$ such that $\mathfrak{a} \subset \mathfrak{p}$. The complement of a closed subset of spec(A) is called an **open subset** of spec(A). The following statements are then very easy to verify, and will be left to the reader.

Proposition 5.2. *The union of a finite number of closed sets is closed. The intersection of an arbitrary family of closed sets is closed.*

The intersection of a finite number of open sets is open. The union of an arbitrary family of open sets is open.

The empty set and spec(A) *itself are both open and closed.*

If S is a subset of A, then the set of prime ideals $\mathfrak{p} \in \mathrm{spec}(A)$ such that $S \subset \mathfrak{p}$ coincides with the set of prime ideals $\mathfrak{p}$ containing the ideal generated by S.

The collection of open sets as in Proposition 5.2 is said to be a **topology** on spec(A), called the **Zariski topology**.

Remark. In analysis, one considers a compact Hausdorff space S. "Hausdorff" means that given two points P, Q there exists disjoint open sets U_P, U_Q containing P and Q respectively. In the present algebraic context, the topology is not Hausdorff. In the analytic context, let R be the ring of complex valued continuous functions on S. Then the maximal ideals of R are in bijection with the points of S (Gelfand-Naimark theorem). To each point $P \in S$, we associate the ideal M_P of functions f such that $f(P) = 0$. The association $P \mapsto M_P$ gives the bijection. There are analogous results in the complex analytic case. For a non-trivial example, see Exercise 19 of Chapter XII.

Let A, B be commutative rings and $\varphi: A \to B$ a homomorphism. Then φ induces a map

$$\varphi^* = \mathrm{spec}(\varphi) = \varphi^{-1} : \mathrm{spec}(B) \to \mathrm{spec}(A)$$

by

$$\mathfrak{p} \mapsto \varphi^{-1}(\mathfrak{p}).$$

Indeed, it is immediately verified that $\varphi^{-1}(\mathfrak{p})$ is a prime ideal of A. Note however that the inverse image of a maximal ideal of B is not necessarily a maximal ideal of A. Example? The reader will verify at once that spec(φ) is continuous, in the sense that if U is open in spec(B), then $\varphi^{-1}(U)$ is open in spec(A).

We can then view spec as a contravariant functor from the category of commutative rings to the category of topological spaces.

By a **point** of spec(A) **in a field** L one means a mapping

$$\mathrm{spec}(\varphi) : \mathrm{spec}(L) \to \mathrm{spec}(A)$$

induced by a homomorphism $\varphi : A \to L$ of A into L.

For example, for each prime number p, we get a point of spec($\mathbf{Z}$), namely the point arising from the reduction map

$$\mathbf{Z} \to \mathbf{Z}/p\mathbf{Z}.$$

The corresponding point is given by the reversed arrow,

$$\operatorname{spec}(\mathbf{Z}) \leftarrow \operatorname{spec}(\mathbf{Z}/p\mathbf{Z}).$$

As another example, consider the polynomial ring $k[X_1, \ldots, X_n]$ over a field k. For each n-tuple $(c_1, \ldots, c_n)$ in $k^{a(n)}$ we get a homomorphism

$$\varphi : k[X_1, \ldots, X_n] \to k^a$$

such that φ is the identity on k, and $\varphi(X_i) = c_i$ for all i. The corresponding point is given by the reversed arrow

$$\operatorname{spec} k[X] \leftarrow \operatorname{spec}(k^a).$$

Thus we may identify the points in n-space $k^{a(n)}$ with the points of spec $k[X]$ (over k) in k^a.

However, one does not want to take points only in the algebraic closure of k, and of course one may deal with the case of an arbitrary variety V over k rather than all of affine n-space. Thus let $k[x_1, \ldots, x_n]$ be a finitely generated entire ring over k with a chosen family of generators. Let $V = \operatorname{spec} k[x]$. Let A be a commutative k-algebra, corresponding to a homomorphism $k \to A$. Then a point of V in A may be described either as a homomorphism

$$\varphi: k[x_1, \ldots, x_n] \to A,$$

or as the reversed arrow

$$\operatorname{spec}(A) \to \operatorname{spec}(k[x])$$

corresponding to this homomorphism. If we put $c_i = \varphi(x_i)$, then one may call $(c) = (c_1, \ldots, c_n)$ the **coordinates of the point in** A. By a **generic point** of V in a field K we mean a point such that the map $\varphi: k[x] \to K$ is injective, i.e. an isomorphism of $k[x]$ with some subring of K.

Let A be a commutative Noetherian ring. We leave it as an exercise to verify the following assertions, which translate the Noetherian condition into properties of closed sets in the Zariski topology.

Closed subsets of spec(A) satisfy the **descending chain condition**, i.e., if

$$C_1 \supset C_2 \supset C_3 \supset \cdots$$

is a descending chain of closed sets, then we have $C_n = C_{n+1}$ for all sufficiently large n. Equivalently, let $\{C_i\}_{i \in I}$ be a family of closed sets. Then there exists a relatively minimal element of this family, that is a closed set C_{i_0} in the family such that for all i, if $C_i \subset C_{i_0}$ then $C_i = C_{i_0}$. The proof follows at once from the corresponding properties of ideals, and the simple formalism relating unions and intersections of closed sets with products and sums of ideals.

A closed set C is said to be **irreducible** if it cannot be expressed as the union of two closed sets

$$C \neq C_1 \cup C_2$$

with $C_1 \neq C$ and $C_2 \neq C$.

Theorem 5.3. *Let A be a Noetherian commutative ring. Then every closed set C can be expressed as a finite union of irreducible closed sets, and this expression is unique if in the union*

$$C = C_1 \cup \cdots \cup C_r$$

of irreducible closed sets, we have $C_i \not\subset C_j$ if $i \neq j$.

Proof. We give the proof as an example to show how the version of Theorem 2.2 has an immediate translation in the more general context of spec(A). Suppose the family of closed sets which cannot be represented as a finite union of irreducible ones is not empty. Translating the Noetherian hypothesis in this case shows that there exists a minimal such set C. Then C cannot be irreducible, and we can write C as a union of closed sets

$$C = C' \cup C'',$$

with $C' \neq C$ and $C'' \neq C$. Since C' and C'' are strictly smaller than C, then we can express C' and C'' as finite unions of irreducible closed sets, thus getting a similar expression for C, and a contradiction which proves existence.

As to uniqueness, let

$$C = C_1 \cup \cdots \cup C_r = Z_1 \cup \cdots \cup Z_s$$

be an expression of C as union of irreducible closed sets, without inclusion relations. For each Z_j we can write

$$Z_j = (Z_j \cap C_1) \cup \cdots \cup (Z_j \cap C_r).$$

Since each $Z_j \cap C_i$ is a closed set, we must have $Z_j = Z_j \cap C_i$ for some i. Hence $Z_j = C_i$ for some i. Similarly, C_i is contained in some Z_k. Since there is no inclusion relation among the Z_j's, we must have $Z_j = C_i = Z_k$. This argument can be carried out for each Z_j and each C_i. This proves that each Z_j appears among the C_i's and each C_i appears among the Z_j's, and proves the uniqueness of our representation. This proves the theorem.

Proposition 5.4. *Let C be a closed subset of* spec(A). *Then C is irreducible if and only if $C = \mathfrak{X}(\mathfrak{p})$ for some prime ideal $\mathfrak{p}$.*

Proof. Exercise.

More properties at the same basic level will be given in Exercises 14–19.

EXERCISES

Integrality

1. (Hilbert-Zariski) Let k be a field and let V be a homogeneous variety with generic point (x) over k. Let $\mathfrak{X}$ be the algebraic set of zeros in k^a of a homogeneous ideal in $k[X]$ generated by forms $f_1, \ldots, f_r$ in $k[X]$. Prove that $V \cap \mathfrak{X}$ has only the trivial zero if and only if each x_i is integral over the ring $k[f(x)] = k[f_1(x), \ldots, f_r(x)]$. (Compare with Theorem 3.7 of Chapter VII.)

2. Let $f_1, \ldots, f_r$ be forms in n variables and suppose $n > r$. Prove that these forms have a non-trivial common zero.

3. Let R be an entire ring. Prove that R is integrally closed if and only if the local ring $R_\mathfrak{p}$ is integrally closed for each prime ideal $\mathfrak{p}$.

4. Let R be an entire ring with quotient field K. Let t be transcendental over K. Let $f(t) = \sum a_i t^i \in K[t]$. Prove:
   (a) If $f(t)$ is integral over $R[t]$, then all a_i are integral over R.
   (b) If R is integrally closed, then $R[t]$ is integrally closed.

For the next exercises, we let $R = k[x] = k[X]/\mathfrak{p}$, where $\mathfrak{p}$ is a homogeneous prime ideal. Then (x) is a homogeneous generic point for a k-variety V. We let I be the integral closure of R in $k(x)$. We assume for simplicity that $k(x)$ is a regular extension of k.

5. Let $z = \sum c_i x_i$ with $c_i \in k$, and $z \neq 0$. If $k[x]$ is integrally closed, prove that $k[x/z]$ is integrally closed.

6. Define an element $f \in k(x)$ to be **homogeneous** if $f(tx) = t^d f(x)$ for t transcendental over $k(x)$ and some integer d. Let $f \in I$. Show that f can be written in the form $f = \sum f_i$ where each f_i is homogeneous of degree $i \geqq 0$, and where also $f_i \in I$. (Some f_i may be 0, of course.)

We let R_m denote the set of elements of R which are homogeneous of degree m. Similarly for I_m. We note that R_m and I_m are vector spaces over k, and that R (resp. I) is the direct sum of all spaces R_m (resp. I_m) for $m = 0, 1, \ldots$ This is obvious for R, and it is true for I because of Exercise 6.

7. Prove that I can be written as a sum $I = Rz_1 + \cdots + Rz_s$, where each z_i is homogeneous of some degree d_i.

8. Define an integer $m \geqq 1$ to be **well behaved** if $I_m^q = I_{qm}$ for all integers $q \geqq 1$. If $R = I$, then all m are well behaved. In Exercise 7, suppose $m \geqq \max d_i$. Show that m is well behaved.

9. (a) Prove that I_m is a finite dimensional vector space over k. Let $w_0, \ldots, w_M$ be a basis for I_m over k. Then $k[I_m] = k[w]$.
   (b) If m is well behaved, show that $k[I_m]$ is integrally closed.
   (c) Denote by $k((x))$ the field generated over k by all quotients x_i/x_j with $x_j \neq 0$, and similarly for $k((w))$. Show that $k((x)) = k((w))$.

(If you want to see Exercises 4–9 worked out, see my *Introduction to Algebraic Geometry*, Interscience 1958, Chapter V.)

Resultants

10. Prove that the resultant defined for n forms in n variables in §3 actually coincides with the resultant of Chapter IV, or §4 when $n = 2$.

11. Let $\mathfrak{a} = (f_1, \ldots, f_r)$ be a homogeneous ideal in $k[X_1, \ldots, X_n)$ (with k algebraically closed). Assume that the only zeros of $\mathfrak{a}$ consist of a finite number of points $(x^{(1)}), \ldots, (x^{(d)})$ in projective space $\mathbf{P}^{n-1}$, so the coordinates of each $x^{(j)}$ can be taken in k. Let $u_1, \ldots, u_n$ be independent variables and let

$$L_u(X) = u_1X_1 + \cdots + u_nX_n.$$

Let $R_1(u), \ldots, R_s(u) \in k[u]$ be a resultant system for $f_1, \ldots, f_r, L_u$.
   (a) Show that the common non-trivial zeros of the system $R_i(u)$ $(i = 1, \ldots, s)$ in k are the zeros of the polynomial

$$\prod_j L_u(x^{(j)}) \in k[u].$$

   (b) Let $D(u)$ be the greatest common divisor of $R_1(u), \ldots, R_s(u)$ in $k[u]$. Show that there exist integers $m_j \geqq 1$ such that (up to a factor in k)

$$D(u) = \prod_{j=1}^{d} L_u(x^{(j)})^{m_j}.$$

[See van der Waerden, *Moderne Algebra*, Second Edition, Volume II, §79.]

12. For forms in 2 variables, prove directly from the definition used in §4 that one has

$$\text{Res}(fg, h) = \text{Res}(f, h)\,\text{Res}(g, h)$$
$$\text{Res}(f, g) = (-1)^{(\deg f)(\deg g)}\text{Res}(g, f).$$

13. Let k be a field and let $\mathbf{Z} \to k$ be the canonical homomorphism. If $F \in \mathbf{Z}[W, X]$, we denote by $\bar{F}$ the image of F in $k[W, X]$ under this homomorphism. Thus we get $\bar{R}$, the image of the resultant R.
   (a) Show that $\bar{R}$ is a generator of the prime ideal $\mathfrak{p}_{k,1}$ of Theorem 3.5 over the field k. Thus we may denote $\bar{R}$ by R_k.
   (b) Show that R is absolutely irreducible, and so is R_k. In other words, R_k is irreducible over the algebraic closure of k.

Spec of a ring

14. Let A be a commutative ring. Define spec(A) to be **connected** if spec(A) is not the union of two disjoint non-empty closed sets (or equivalently, spec(A) is not the union of two disjoint, non-empty open sets).
   (a) Suppose that there are idempotents e_1, e_2 in A (that is $e_1^2 = e_1$ and $e_2^2 = e_2$), $\neq 0, 1$, such that $e_1e_2 = 0$ and $e_1 + e_2 = 1$. Show that spec(A) is not connected.
   (b) Conversely, if spec(A) is not connected, show that there exist idempotents as in part (a).

In either case, the existence of the idempotents is equivalent with the fact that the ring A is a product of two non-zero rings, $A = A_1 \times A_2$.

15. Prove that the Zariski topology is **compact**, in other words: let $\{U_i\}_{i \in I}$ be a family of open sets such that

$$\bigcup_i U_i = \operatorname{spec}(A).$$

Show that there is a finite number of open sets $U_{i_1}, \ldots, U_{i_n}$ whose union is spec(A). [*Hint*: Use closed sets, and use the fact that if a sum of ideals is the unit ideal, then 1 can be written as a finite sum of elements.]

16. Let f be an element of A. Let S be the multiplicative subset $\{1, f, f^2, f^3, \ldots\}$ consisting of the powers of f. We denote by A_f the ring $S^{-1}A$ as in Chapter II, §3. From the natural homomorphism $A \to A_f$ one gets the corresponding map spec$(A_f) \to$ spec(A).
    (a) Show that spec(A_f) maps on the open set of points in spec(A) which are not zeros of f.
    (b) Given a point $\mathfrak{p} \in$ spec(A), and an open set U containing $\mathfrak{p}$, show that there exists f such that $\mathfrak{p} \in$ spec$(A_f) \subset U$.

17. Let U_i = spec(A_{f_i}) be a finite family of open subsets of spec(A) covering spec(A). For each i, let $a_i/f_i \in A_{f_i}$. Assume that as functions on $U_i \cap U_j$ we have $a_i/f_i = a_j/f_j$ for all pairs i, j. Show that there exists a unique element $a \in A$ such that $a = a_i/f_i$ in A_{f_i} for all i.

18. Let k be a field and let $k[x_1, \ldots, x_n] = A \subset K$ be a finitely generated subring of some extension field K. Assume that $k(x_1, \ldots, x_n)$ has transcendence degree r. Show that every maximal chain of prime ideals

$$A \supset P_1 \supset P_2 \supset \ldots \supset P_m \supset \{0\},$$

with $P_1 \neq A$, $P_i \neq P_{i+1}$, $P_m \neq \{0\}$, must have $m = r$.

19. Let $A = \mathbf{Z}[x_1, \ldots, x_n]$ be a finitely generated entire ring over $\mathbf{Z}$. Show that every maximal chain of prime ideals as in Exercise 18 must have $m = r + 1$. Here, r = transcendence degree of $\mathbf{Q}(x_1, \ldots, x_n)$ over $\mathbf{Q}$.

CHAPTER X

Noetherian Rings and Modules

This chapter may serve as an introduction to the methods of algebraic geometry rooted in commutative algebra and the theory of modules, mostly over a Noetherian ring.

§1. BASIC CRITERIA

Let A be a ring and M a module (i.e., a left A-module). We shall say that M is **Noetherian** if it satisfies any one of the following three conditions:

(1) Every submodule of M is finitely generated.

(2) Every ascending sequence of submodules of M,

$$M_1 \subset M_2 \subset M_3 \subset \cdots,$$

such that $M_i \neq M_{i+1}$ is finite.

(3) Every non-empty set S of submodules of M has a maximal element (i.e., a submodule M_0 such that for any element N of S which contains M_0 we have $N = M_0$).

We shall now prove that the above three conditions are equivalent.

(1) $\Rightarrow$ (2) Suppose we have an ascending sequence of submodules of M as above. Let N be the union of all the M_i ($i = 1, 2, \ldots$). Then N is finitely generated, say by elements $x_1, \ldots, x_r$, and each generator is in some M_i. Hence there exists an index j such that

$$x_1, \ldots, x_r \in M_j.$$

Then

$$\langle x_1, \ldots, x_r \rangle \subset M_j \subset N = \langle x_1, \ldots, x_r \rangle,$$

whence equality holds and our implication is proved.

(2) $\Rightarrow$ (3) Let N_0 be an element of S. If N_0 is not maximal, it is properly contained in a submodule N_1. If N_1 is not maximal, it is properly contained in a submodule N_2. Inductively, if we have found N_i which is not maximal, it is contained properly in a submodule N_{i+1}. In this way we could construct an infinite chain, which is impossible.

(3) $\Rightarrow$ (1) Let N be a submodule of M. Let $a_0 \in N$. If $N \neq \langle a_0 \rangle$, then there exists an element $a_1 \in N$ which does not lie in $\langle a_0 \rangle$. Proceeding inductively, we can find an ascending sequence of submodules of N, namely

$$\langle a_0 \rangle \subset \langle a_0, a_1 \rangle \subset \langle a_0, a_1, a_2 \rangle \subset \cdots$$

where the inclusion each time is proper. The set of these submodules has a maximal element, say a submodule $\langle a_0, a_1, \ldots, a_r \rangle$, and it is then clear that this finitely generated submodule must be equal to N, as was to be shown.

Proposition 1.1. *Let M be a Noetherian A-module. Then every submodule and every factor module of M is Noetherian.*

Proof. Our assertion is clear for submodules (say from the first condition). For the factor module, let N be a submodule and $f: M \to M/N$ the canonical homomorphism. Let $\overline{M}_1 \subset \overline{M}_2 \subset \cdots$ be an ascending chain of submodules of M/N and let $M_i = f^{-1}(\overline{M}_i)$. Then $M_1 \subset M_2 \subset \cdots$ is an ascending chain of submodules of M, which must have a maximal element, say M_r, so that $M_i = M_r$ for $r \geqq i$. Then $f(M_i) = \overline{M}_i$ and our assertion follows.

Proposition 1.2. *Let M be a module, N a submodule. Assume that N and M/N are Noetherian. Then M is Noetherian.*

Proof. With every submodule L of M we associate the pair of modules

$$L \mapsto (L \cap N, (L + N)/N).$$

We contend: If $E \subset F$ are two submodules of M such that their associated pairs are equal, then $E = F$. To see this, let $x \in F$. By the hypothesis that $(E + N)/N = (F + N)/N$ there exist elements $u, v \in N$ and $y \in E$ such that $y + u = x + v$. Then

$$x - y = u - v \in F \cap N = E \cap N.$$

Since $y \in E$, it follows the $x \in E$ and our contention is proved. If we have an ascending sequence

$$E_1 \subset E_2 \subset \cdots$$

then the associated pairs form an ascending sequence of submodules of N and M/N respectively, and these sequences must stop. Hence our sequence $E_1 \subset E_2 \cdots$ also stops, by our preceding contention.

Propositions 1.1 and 1.2 may be summarized by saying that in an exact sequence $0 \to M' \to M \to M'' \to 0$, M is Noetherian if and only if M' and M'' are Noetherian.

Corollary 1.3. *Let M be a module, and let N, N' be submodules. If $M = N + N'$ and if both N, N' are Noetherian, then M is Noetherian. A finite direct sum of Noetherian modules is Noetherian.*

Proof. We first observe that the direct product $N \times N'$ is Noetherian since it contains N as a submodule whose factor module is isomorphic to N', and Proposition 1.2 applies. We have a surjective homomorphism

$$N \times N' \to M$$

such that the pair (x, x') with $x \in N$ and $x' \in N'$ maps on $x + x'$. By Proposition 1.1, it follows that M is Noetherian. Finite products (or sums) follow by induction.

A ring A is called **Noetherian** if it is Noetherian as a left module over itself. This means that every left ideal is finitely generated.

Proposition 1.4. *Let A be a Noetherian ring and let M be a finitely generated module. Then M is Noetherian.*

Proof. Let $x_1, \ldots, x_n$ be generators of M. There exists a homomorphism

$$f: A \times A \times \cdots \times A \to M$$

of the product of A with itself n times such that

$$f(a_1, \ldots, a_n) = a_1 x_1 + \cdots + a_n x_n.$$

This homomorphism is surjective. By the corollary of the preceding proposition, the product is Noetherian, and hence M is Noetherian by Proposition 1.1.

Proposition 1.5. *Let A be a ring which is Noetherian, and let $\varphi: A \to B$ be a surjective ring-homomorphism. Then B is Noetherian.*

Proof. Let $\mathfrak{b}_1 \subset \cdots \subset \mathfrak{b}_n \subset \cdots$ be an ascending chain of left ideals of B and let $\mathfrak{a}_i = \varphi^{-1}(\mathfrak{b}_i)$. Then the $\mathfrak{a}_i$ form an ascending chain of left ideals of A which must stop, say at $\mathfrak{a}_r$. Since $\varphi(\mathfrak{a}_i) = \mathfrak{b}_i$ for all i, our proposition is proved.

Proposition 1.6. *Let A be a commutative Noetherian ring, and let S be a multiplicative subset of A. Then $S^{-1}A$ is Noetherian.*

Proof. We leave the proof as an exercise.

Examples. In Chapter IV, we gave the fundamental examples of Noetherian rings, namely polynomial rings and rings of power series. The above propositions show how to construct other examples from these, by taking factor rings or modules, or submodules.

We have already mentioned that for applications to algebraic geometry, it is valuable to consider factor rings of type $k[X]/\mathfrak{a}$, where $\mathfrak{a}$ is an arbitrary ideal. For this and similar reasons, it has been found that the foundations should be laid in terms of modules, not just ideals or factor rings. Notably, we shall first see that the prime ideal associated with an irreducible algebraic set has an analogue in terms of modules. We shall also see that the decomposition of an algebraic set into irreducibles has a natural formulation in terms of modules, namely by expressing a submodule as an intersection or primary modules.

In §6 we shall apply some general notions to get the Hilbert polynomial of a module of finite length, and we shall make comments on how this can be interpreted in terms of geometric notions. Thus the present chapter is partly intended to provide a bridge between basic algebra and algebraic geometry.

§2. ASSOCIATED PRIMES

Throughout this section, we let A be a commutative ring. Modules and homomorphisms are A-modules and A-homomorphisms unless otherwise specified.

Proposition 2.1. *Let S be a multiplicative subset of A, and assume that S does not contain 0. Then there exists an ideal of A which is maximal in the set of ideals not intersecting S, and any such ideal is prime.*

Proof. The existence of such an ideal $\mathfrak{p}$ follows from Zorn's lemma (the set of ideals not meeting S is not empty, because it contains the zero ideal, and is clearly inductively ordered). Let $\mathfrak{p}$ be maximal in the set. Let $a, b \in A$, $ab \in \mathfrak{p}$, but $a \notin \mathfrak{p}$ and $b \notin \mathfrak{p}$. By hypothesis, the ideals $(a, \mathfrak{p})$ and $(b, \mathfrak{p})$ generated by a and $\mathfrak{p}$ (or b and $\mathfrak{p}$ respectively) meet S, and there exist therefore elements $s, s' \in S$, $c, c', x, x' \in A$, $p, p' \in \mathfrak{p}$ such that

$$s = ca + xp \quad \text{and} \quad s' = c'b + x'p'.$$

Multiplying these two expressions, we obtain

$$ss' = cc'ab + p''$$

with some $p'' \in \mathfrak{p}$, whence we see that ss' lies in $\mathfrak{p}$. This contradicts the fact that $\mathfrak{p}$ does not intersect S, and proves that $\mathfrak{p}$ is prime.

An element a of A is said to be **nilpotent** if there exists an integer $n \geqq 1$ such that $a^n = 0$.

Corollary 2.2. *An element a of A is nilpotent if and only if it lies in every prime ideal of A.*

Proof. If $a^n = 0$, then $a^n \in \mathfrak{p}$ for every prime $\mathfrak{p}$, and hence $a \in \mathfrak{p}$. If $a^n \neq 0$ for any positive integer n, we let S be the multiplicative subset of powers of a, namely $\{1, a, a^2, \dots\}$, and find a prime ideal as in the proposition to prove the converse.

Let $\mathfrak{a}$ be an ideal of A. The **radical** of $\mathfrak{a}$ is the set of all $a \in A$ such that $a^n \in \mathfrak{a}$ for some integer $n \geqq 1$, (or equivalently, it is the set of elements $a \in A$ whose image in the factor ring $A/\mathfrak{a}$ is nilpotent). We observe that the radical of $\mathfrak{a}$ is an ideal, for if $a^n = 0$ and $b^m = 0$ then $(a + b)^k = 0$ if k is sufficiently large: In the binomial expansion, either a or b will appear with a power at least equal to n or m.

Corollary 2.3. *An element a of A lies in the radical of an ideal $\mathfrak{a}$ if and only if it lies in every prime ideal containing $\mathfrak{a}$.*

Proof. Corollary 2.3 is equivalent to Corollary 2.2 applied to the ring $A/\mathfrak{a}$.

We shall extend Corollary 2.2 to modules. We first make some remarks on localization. Let S be a multiplicative subset of A. If M is a module, we can define $S^{-1}M$ in the same way that we defined $S^{-1}A$. We consider equivalence classes of pairs (x, s) with $x \in M$ and $s \in S$, two pairs (x, s) and (x', s') being equivalent if there exists $s_1 \in S$ such that $s_1(s'x - sx') = 0$. We denote the equivalence class of (x, s) by x/s, and verify at once that the set of equivalence classes is an additive group (under the obvious operations). It is in fact an A-module, under the operation

$$(a, x/s) \mapsto ax/s.$$

We shall denote this module of equivalence classes by $S^{-1}M$. (We note that $S^{-1}M$ could also be viewed as an $S^{-1}A$-module.)

If $\mathfrak{p}$ is a prime ideal of A, and S is the complement of $\mathfrak{p}$ in A, then $S^{-1}M$ is also denoted by $M_\mathfrak{p}$.

It follows trivially from the definitions that if $N \to M$ is an injective homomorphism, then we have a natural injection $S^{-1}N \to S^{-1}M$. In other words, if N is a submodule of M, then $S^{-1}N$ can be viewed as a submodule of $S^{-1}M$. If $x \in N$ and $s \in S$, then the fraction x/s can be viewed as an element of $S^{-1}N$ or $S^{-1}M$. If $x/s = 0$ in $S^{-1}M$, then there exists $s_1 \in S$ such that $s_1 x = 0$, and this means that x/s is also 0 in $S^{-1}N$. Thus if $\mathfrak{p}$ is a prime ideal and N is a submodule of M, we have a natural inclusion of $N_\mathfrak{p}$ in $M_\mathfrak{p}$. We shall in fact identify $N_\mathfrak{p}$ as a submodule of $M_\mathfrak{p}$. In particular, we see that $M_\mathfrak{p}$ is the sum of its submodules $(Ax)_\mathfrak{p}$, for $x \in M$ (but of course not the direct sum).

Let $x \in M$. The **annihilator** $\mathfrak{a}$ of x is the ideal consisting of all elements $a \in A$ such that $ax = 0$. We have an isomorphism (of modules)

$$A/\mathfrak{a} \xrightarrow{\approx} Ax$$

under the map

$$a \to ax.$$

Lemma 2.4. *Let x be an element of a module M, and let $\mathfrak{a}$ be its annihilator. Let $\mathfrak{p}$ be a prime ideal of A. Then $(Ax)_{\mathfrak{p}} \neq 0$ if and only if $\mathfrak{p}$ contains $\mathfrak{a}$.*

Proof. The lemma is an immediate consequence of the definitions, and will be left to the reader.

Let a be an element of A. Let M be a module. The homomorphism

$$x \mapsto ax, \qquad x \in M$$

will be called the **principal homomorphism** associated with a, and will be denoted by a_M. We shall say that a_M is **locally nilpotent** if for each $x \in M$ there exists an integer $n(x) \geqq 1$ such that $a^{n(x)}x = 0$. This condition implies that for every finitely generated submodule N of M, there exists an integer $n \geqq 1$ such that $a^n N = 0$: We take for n the largest power of a annihilating a finite set of generators of N. Therefore, *if M is finitely generated, a_M is locally nilpotent if and only if it is nilpotent.*

Proposition 2.5. *Let M be a module, $a \in A$. Then a_M is locally nilpotent if and only if a lies in every prime ideal $\mathfrak{p}$ such that $M_{\mathfrak{p}} \neq 0$.*

Proof. Assume that a_M is locally nilpotent. Let $\mathfrak{p}$ be a prime of A such that $M_{\mathfrak{p}} \neq 0$. Then there exists $x \in M$ such that $(Ax)_{\mathfrak{p}} \neq 0$. Let n be a positive integer such that $a^n x = 0$. Let $\mathfrak{a}$ be the annihilator of x. Then $a^n \in \mathfrak{a}$, and hence we can apply the lemma, and Corollary 4.3 to conclude that a lies in every prime $\mathfrak{p}$ such that $M_{\mathfrak{p}} \neq 0$. Conversely, suppose a_M is not locally nilpotent, so there exists $x \in M$ such that $a^n x = 0$ for all $n \geqq 0$. Let $S = \{1, a, a^2, \ldots\}$, and using Proposition 2.1 let $\mathfrak{p}$ be a prime not intersecting S. Then $(Ax)_{\mathfrak{p}} \neq 0$, so $M_{\mathfrak{p}} \neq 0$ and $a \notin \mathfrak{p}$, as desired.

Let M be a module. A prime ideal $\mathfrak{p}$ of A will be said to be **associated** with M if there exists an element $x \in M$ such that $\mathfrak{p}$ is the annihilator of x. In particular, since $\mathfrak{p} \neq A$, we must have $x \neq 0$.

Proposition 2.6. *Let M be a module $\neq 0$. Let $\mathfrak{p}$ be a maximal element in the set of ideals which are annihilators of elements $x \in M$, $x \neq 0$. Then $\mathfrak{p}$ is prime.*

Proof. Let $\mathfrak{p}$ be the annihilator of the element $x \neq 0$. Then $\mathfrak{p} \neq A$. Let $a, b \in A$, $ab \in \mathfrak{p}$, $a \notin \mathfrak{p}$. Then $ax \neq 0$. But the ideal $(b, \mathfrak{p})$ annihilates ax, and contains $\mathfrak{p}$. Since $\mathfrak{p}$ is maximal, it follows that $b \in \mathfrak{p}$, and hence $\mathfrak{p}$ is prime.

Corollary 2.7. *If A is Noetherian and M is a module $\neq 0$, then there exists a prime associated with M.*

Proof. The set of ideals as in Proposition 2.6 is not empty since $M \neq 0$, and has a maximal element because A is Noetherian.

Corollary 2.8. *Assume that both A and M are Noetherian, $M \neq 0$. Then there exists a sequence of submodules*

$$M = M_1 \supset M_2 \supset \cdots \supset M_r = 0$$

such that each factor module M_i/M_{i+1} is isomorphic to $A/\mathfrak{p}_i$ for some prime $\mathfrak{p}_i$.

Proof. Consider the set of submodules having the property described in the corollary. It is not empty, since there exists an associated prime $\mathfrak{p}$ of M, and if $\mathfrak{p}$ is the annihilator of x, then $Ax \approx A/\mathfrak{p}$. Let N be a maximal element in the set. If $N \neq M$, then by the preceding argument applied to M/N, there exists a submodule N' of M containing N such that N'/N is isomorphic to $A/\mathfrak{p}$ for some $\mathfrak{p}$, and this contradicts the maximality of N.

Proposition 2.9. *Let A be Noetherian, and $a \in A$. Let M be a module. Then a_M is injective if and only if a does not lie in any associated prime of M.*

Proof. Assume that a_M is not injective, so that $ax = 0$ for some $x \in M$, $x \neq 0$. By Corollary 2.7, there exists an associated prime $\mathfrak{p}$ of Ax, and a is an element of $\mathfrak{p}$. Conversely, if a_M is injective, then a cannot lie in any associated prime because a does not annihilate any non-zero element of M.

Proposition 2.10. *Let A be Noetherian, and let M be a module. Let $a \in A$. The following conditions are equivalent:*

(i) *a_M is locally nilpotent.*

(ii) *a lies in every associated prime of M.*

(iii) *a lies in every prime $\mathfrak{p}$ such that $M_\mathfrak{p} \neq 0$.*

If $\mathfrak{p}$ is a prime such that $M_\mathfrak{p} \neq 0$, then $\mathfrak{p}$ contains an associated prime of M.

Proof. The fact that (i) implies (ii) is obvious from the definitions, and does not need the hypothesis that A is Noetherian. Neither does the fact that (iii) implies (i), which has been proved in Proposition 2.5. We must therefore prove that (ii) implies (iii) which is actually implied by the last statement. The latter is proved as follows. Let $\mathfrak{p}$ be a prime such that $M_\mathfrak{p} \neq 0$. Then there exists $x \in M$ such that $(Ax)_\mathfrak{p} \neq 0$. By Corollary 2.7, there exists an associated prime $\mathfrak{q}$ of $(Ax)_\mathfrak{p}$ in A. Hence there exists an element y/s of $(Ax)_\mathfrak{p}$, with $y \in Ax$, $s \notin \mathfrak{p}$, and $y/s \neq 0$, such that $\mathfrak{q}$ is the annihilator of y/s. It follows that $\mathfrak{q} \subset \mathfrak{p}$, for otherwise, there exists $b \in \mathfrak{q}$, $b \notin \mathfrak{p}$, and $0 = by/s$, whence $y/s = 0$, contradiction. Let $b_1, \ldots, b_n$ be generators for $\mathfrak{q}$. For each i, there exists $s_i \in A$, $s_i \notin \mathfrak{p}$, such that $s_i b_i y = 0$ because $b_i y/s = 0$. Let $t = s_1 \cdots s_n$. Then it is trivially verified that $\mathfrak{q}$ is the annihilator of ty in A. Hence $\mathfrak{q} \subset \mathfrak{p}$, as desired.

Let us define the **support** of M by

$\operatorname{supp}(M) =$ set of primes $\mathfrak{p}$ such that $M_\mathfrak{p} \neq 0$.

We also have the **annihilator** of M,

$$\operatorname{ann}(M) = \text{set of elements } a \in A \text{ such that } aM = 0.$$

We use the notation

$$\operatorname{ass}(M) = \text{set of associated primes of } M.$$

For any ideal $\mathfrak{a}$ we have its **radical**,

$$\operatorname{rad}(\mathfrak{a}) = \text{set of elements } a \in A \text{ such that } a^n \in \mathfrak{a} \text{ for some integer } n \geqq 1.$$

Then for *finitely generated M*, we can reformulate Proposition 2.10 by the following formula:

$$\boxed{\operatorname{rad}(\operatorname{ann}(M)) = \bigcap_{\mathfrak{p} \in \operatorname{supp}(M)} \mathfrak{p} = \bigcap_{\mathfrak{p} \in \operatorname{ass}(M)} \mathfrak{p}.}$$

Corollary 2.11. *Let A be Noetherian, and let M be a module. The following conditions are equivalent:*

(i) *There exists only one associated prime of M.*

(ii) *We have $M \neq 0$, and for every $a \in A$, the homomorphism a_M is injective, or locally nilpotent.*

If these conditions are satisfied, then the set of elements $a \in A$ such that a_M is locally nilpotent is equal to the associated prime of M.

Proof. Immediate consequence of Propositions 2.9 and 2.10.

Proposition 2.12. *Let N be a submodule of M. Every associated prime of N is associated with M also. An associated prime of M is associated with N or with M/N.*

Proof. The first assertion is obvious. Let $\mathfrak{p}$ be an associated prime of M, and say $\mathfrak{p}$ is the annihilator of the element $x \neq 0$. If $Ax \cap N = 0$, then Ax is isomorphic to a submodule of M/N, and hence $\mathfrak{p}$ is associated with M/N. Suppose $Ax \cap N \neq 0$. Let $y = ax \in N$ with $a \in A$ and $y \neq 0$. Then $\mathfrak{p}$ annihilates y. We claim $\mathfrak{p} = \operatorname{ann}(y)$. Let $b \in A$ and $by = 0$. Then $ba \in \mathfrak{p}$ but $a \notin \mathfrak{p}$, so $b \in \mathfrak{p}$. Hence $\mathfrak{p}$ is the annihilator of y in A, and therefore is associated with N, as was to be shown.

§3. PRIMARY DECOMPOSITION

We continue to assume that A is a commutative ring, and that modules (resp. homomorphisms) are A-modules (resp. A-homomorphisms), unless otherwise specified.

Let M be a module. A submodule Q of M is said to be **primary** if $Q \neq M$, and if given $a \in A$, the homomorphism $a_{M/Q}$ is either injective or nilpotent. Viewing A as a module over itself, we see that an ideal $\mathfrak{q}$ is **primary** if and only if it satisfies the following condition:

Given $a, b \in A$, $ab \in \mathfrak{q}$ and $a \notin \mathfrak{q}$, then $b^n \in \mathfrak{q}$ for some $n \geqq 1$.

Let Q be primary. Let $\mathfrak{p}$ be the ideal of elements $a \in A$ such that $a_{M/Q}$ is nilpotent. Then $\mathfrak{p}$ is prime. Indeed, suppose that $a, b \in A$, $ab \in \mathfrak{p}$ and $a \notin \mathfrak{p}$. Then $a_{M/Q}$ is injective, and consequently $a^n_{M/Q}$ is injective for all $n \geqq 1$. Since $(ab)_{M/Q}$ is nilpotent, it follows that $b_{M/Q}$ must be nilpotent, and hence that $b \in \mathfrak{p}$, proving that $\mathfrak{p}$ is prime. We shall call $\mathfrak{p}$ the prime **belonging** to Q, and also say that Q is $\mathfrak{p}$-primary.

We note the corresponding property for a primary module Q with prime $\mathfrak{p}$:

Let $b \in A$ and $x \in M$ be such that $bx \in Q$. If $x \notin Q$ then $b \in \mathfrak{p}$.

Examples. Let $\mathfrak{m}$ be a maximal ideal of A and let $\mathfrak{q}$ be an ideal of A such that $\mathfrak{m}^k \subset \mathfrak{q}$ for some positive integer k. Then $\mathfrak{q}$ is primary, and $\mathfrak{m}$ belongs to $\mathfrak{q}$. We leave the proof to the reader.

The above conclusion is not always true if $\mathfrak{m}$ is replaced by some prime ideal $\mathfrak{p}$. For instance, let R be a factorial ring with a prime element t. Let A be the subring of polynomials $f(X) \in R[X]$ such that

$$f(X) = a_0 + a_1X + \dots$$

with a_1 divisible by t. Let $\mathfrak{p} = (tX, X^2)$. Then $\mathfrak{p}$ is prime but

$$\mathfrak{p}^2 = (t^2X^2, tX^3, X^4)$$

is not primary, as one sees because $X^2 \notin \mathfrak{p}^2$ but $t^k \notin \mathfrak{p}^2$ for all $k \geqq 1$, yet $t^2X^2 \in \mathfrak{p}^2$.

Proposition 3.1. *Let M be a module, and $Q_1, \dots, Q_r$ submodules which are $\mathfrak{p}$-primary for the same prime $\mathfrak{p}$. Then $Q_1 \cap \dots \cap Q_r$ is also $\mathfrak{p}$-primary.*

Proof. Let $Q = Q_1 \cap \dots \cap Q_r$. Let $a \in \mathfrak{p}$. Let n_i be such that $(a_{M/Q_i})^{n_i} = 0$ for each $i = 1, \dots, r$ and let n be the maximum of $n_1, \dots, n_r$. Then $a^n_{M/Q} = 0$, so that $a_{M/Q}$ is nilpotent. Conversely, suppose $a \notin \mathfrak{p}$. Let $x \in M$, $x \notin Q_j$ for some j. Then $a^n x \notin Q_j$ for all positive integers n, and consequently $a_{M/Q}$ is injective. This proves our proposition.

Let N be a submodule of M. When N is written as a finite intersection of primary submodules, say

$$N = Q_1 \cap \cdots \cap Q_r,$$

we shall call this a **primary decomposition** of N. Using Proposition 3.1, we see that by grouping the Q_i according to their primes, we can always obtain from a given primary decomposition another one such that the primes belonging to the primary ideals are all distinct. A primary decomposition as above such that the prime ideals $\mathfrak{p}_1, \ldots, \mathfrak{p}_r$ belonging to $Q_1, \ldots, Q_r$ respectively are distinct, and such that N cannot be expressed as an intersection of a proper subfamily of the primary ideals $\{Q_1, \ldots, Q_r\}$ will be said to be **reduced**. By deleting some of the primary modules appearing in a given decomposition, we see that if N admits some primary decomposition, then it admits a reduced one. We shall prove a result giving certain uniqueness properties of a reduced primary decomposition.

Let N be a submodule of M and let $x \mapsto \bar{x}$ be the canonical homomorphism. Let $\overline{Q}$ be a submodule of $\overline{M} = M/N$ and let Q be its inverse image in M. Then directly from the definition, one sees that $\overline{Q}$ is primary if and only if Q is primary; and if they are primary, then the prime belonging to Q is also the prime belonging to $\overline{Q}$. Furthermore, if $N = Q_1 \cap \ldots \cap Q_r$ is a primary decomposition of N in M, then

$$(0) = \overline{Q}_1 \cap \ldots \cap \overline{Q}_r$$

is a primary decomposition of (0) in $\overline{M}$, as the reader will verify at once from the definitions. In addition, the decomposition of N is reduced if and only if the decomposition of (0) is reduced since the primes belonging to one are the same as the primes belonging to the other.

Let $Q_1 \cap \cdots \cap Q_r = N$ be a reduced primary decomposition, and let $\mathfrak{p}_i$ belong to Q_i. If $\mathfrak{p}_i$ does not contain $\mathfrak{p}_j$ $(j \neq i)$ then we say that $\mathfrak{p}_i$ is **isolated**. The isolated primes are therefore those primes which are minimal in the set of primes belonging to the primary modules Q_i.

Theorem 3.2. *Let N be a submodule of M, and let*

$$N = Q_1 \cap \cdots \cap Q_r = Q'_1 \cap \cdots \cap Q'_s$$

be a reduced primary decomposition of N. Then $r = s$. The set of primes belonging to $Q_1, \ldots, Q_r$ and $Q'_1, \ldots, Q'_r$ is the same. If $\{\mathfrak{p}_1, \ldots, \mathfrak{p}_m\}$ is the set of isolated primes belonging to these decompositions, then $Q_i = Q'_i$ for $i = 1, \ldots, m$, in other words, the primary modules corresponding to isolated primes are uniquely determined.

Proof. The uniqueness of the number of terms in a reduced decomposition and the uniqueness of the family of primes belonging to the primary components will be a consequence of Theorem 3.5 below.

There remains to prove the uniqueness of the primary module belonging to an isolated prime, say $\mathfrak{p}_1$. By definition, for each $j = 2, \ldots, r$ there exists $a_j \in \mathfrak{p}_j$ and $a_j \notin \mathfrak{p}_1$. Let $a = a_2 \cdots a_r$ be the product. Then $a \in \mathfrak{p}_j$ for all $j > 1$, but $a \notin \mathfrak{p}_1$. We can find an integer $n \geqq 1$ such that $a^n_{M/Q_j} = 0$ for $j = 2, \ldots, r$. Let

$$N_1 = \text{set of } x \in M \text{ such that } a^n x \in N.$$

We contend that $Q_1 = N_1$. This will prove the desired uniqueness. Let $x \in Q_1$. Then $a^n x \in Q_1 \cap \cdots \cap Q_r = N$, so $x \in N_1$. Conversely, let $x \in N_1$, so that $a^n x \in N$, and in particular $a^n x \in Q_1$. Since $a \notin \mathfrak{p}_1$, we know by definition that a_{M/Q_1} is injective. Hence $x \in Q_1$, thereby proving our theorem.

Theorem 3.3. *Let M be a Noetherian module. Let N be a submodule of M. Then N admits a primary decomposition.*

Proof. We consider the set of submodules of M which do not admit a primary decomposition. If this set is not empty, then it has a maximal element because M is Noetherian. Let N be this maximal element. Then N is not primary, and there exists $a \in A$ such that $a_{M/N}$ is neither injective nor nilpotent. The increasing sequence of modules

$$\operatorname{Ker} a_{M/N} \subset \operatorname{Ker} a^2_{M/N} \subset \operatorname{Ker} a^3_{M/N} \subset \cdots$$

stops, say at $a^r_{M/N}$. Let $\varphi : M/N \to M/N$ be the endomorphism $\varphi = a^r_{M/N}$. Then $\operatorname{Ker} \varphi^2 = \operatorname{Ker} \varphi$. Hence $0 = \operatorname{Ker} \varphi \cap \operatorname{Im} \varphi$ in M/N, and neither the kernel nor the image of φ is 0. Taking the inverse image in M, we see that N is the intersection of two submodules of M, unequal to N. We conclude from the maximality of N that each one of these submodules admits a primary decomposition, and therefore that N admits one also, contradiction.

We shall conclude our discussion by relating the primes belonging to a primary decomposition with the associated primes discussed in the previous section.

Proposition 3.4. *Let A and M be Noetherian. A submodule Q of M is primary if and only if M/Q has exactly one associated prime $\mathfrak{p}$, and in that case, $\mathfrak{p}$ belongs to Q, i.e. Q is $\mathfrak{p}$-primary.*

Proof. Immediate consequence of the definitions, and Corollary 2.11.

Theorem 3.5. *Let A and M be Noetherian. The associated primes of M are precisely the primes which belong to the primary modules in a reduced primary decomposition of 0 in M. In particular, the set of associated primes of M is finite.*

Proof. Let

$$0 = Q_1 \cap \cdots \cap Q_r$$

be a reduced primary decomposition of 0 in M. We have an injective homomorphism

$$M \to \bigoplus_{i=1}^{r} M/Q_i.$$

By Proposition 2.12 and Proposition 3.4, we conclude that every associated prime of M belongs to some Q_i. Conversely, let $N = Q_2 \cap \cdots \cap Q_r$. Then $N \neq 0$ because our decomposition is reduced. We have

$$N = N/(N \cap Q_1) \approx (N + Q_1)/Q_1 \subset M/Q_1.$$

Hence N is isomorphic to a submodule of M/Q_1, and consequently has an associated prime which can be none other than the prime $\mathfrak{p}_1$ belonging to Q_1. This proves our theorem.

Theorem 3.6. *Let A be a Noetherian ring. Then the set of divisors of zero in A is the set-theoretic union of all primes belonging to primary ideals in a reduced primary decomposition of* 0.

Proof. An element of $a \in A$ is a divisor of 0 if and only if a_A is not injective. According to Proposition 2.9, this is equivalent to a lying in some associated prime of A (viewed as module over itself). Applying Theorem 3.5 concludes the proof.

§4. NAKAYAMA'S LEMMA

We let A denote a commutative ring, but not necessarily Noetherian.

When dealing with modules over a ring, many properties can be obtained first by localizing, thus reducing problems to modules over local rings. In practice, as in the present section, such modules will be finitely generated. This section shows that some aspects can be reduced to vector spaces over a field by reducing modulo the maximal ideal of the local ring. Over a field, a module always has a basis. We extend this property as far as we can to modules finite over a local ring. The first three statements which follow are known as **Nakayama's lemma**.

Lemma 4.1. *Let $\mathfrak{a}$ be an ideal of A which is contained in every maximal ideal of A. Let E be a finitely generated A-module. Suppose that $\mathfrak{a}E = E$. Then $E = \{0\}$.*

Proof. Induction on the number of generators of E. Let $x_1, \ldots, x_s$ be generators of E. By hypothesis, there exist elements $a_1, \ldots, a_s \in \mathfrak{a}$ such that

$$x_s = a_1 x_1 + \cdots + a_s x_s,$$

so there is an element a (namely a_s) in $\mathfrak{a}$ such that $(1 + a)x_s$ lies in the module generated by the first $s - 1$ generators. Furthermore $1 + a$ is a unit in A, otherwise $1 + a$ is contained in some maximal ideal, and since a lies in all maximal ideals, we conclude that 1 lies in a maximal ideal, which is not possible. Hence x_s itself lies in the module generated by $s - 1$ generators, and the proof is complete by induction.

Lemma 4.1 applies in particular to the case when A is a local ring, and $\mathfrak{a} = \mathfrak{m}$ is its maximal ideal.

Lemma 4.2. *Let A be a local ring, let E be a finitely generated A-module, and F a submodule. If $E = F + \mathfrak{m}E$, then $E = F$.*

Proof. Apply Lemma 4.1 to E/F.

Lemma 4.3. *Let A be a local ring. Let E be a finitely generated A-module. If $x_1, \ldots, x_n$ are generators for E* mod $\mathfrak{m}E$, *then they are generators for E.*

Proof. Take F to be the submodule generated by $x_1, \ldots, x_n$.

Theorem 4.4. *Let A be a local ring and E a finite projective A-module. Then E is free. In fact, if $x_1, \ldots, x_n$ are elements of E whose residue classes $\bar{x}_1, \ldots, \bar{x}_n$ are a basis of $E/\mathfrak{m}E$ over $A/\mathfrak{m}$, then $x_1, \ldots, x_n$ are a basis of E over A. If $x_1, \ldots, x_r$ are such that $\bar{x}_1, \ldots, \bar{x}_r$ are linearly independent over $A/\mathfrak{m}$, then they can be completed to a basis of E over A.*

Proof. I am indebted to George Bergman for the following proof of the first statement. Let F be a free module with basis $e_1, \ldots, e_n$, and let $f\colon F \to E$ be the homomorphism mapping e_i to x_i. We want to prove that f is an isomorphism. By Lemma 4.3, f is surjective. Since E is projective, it follows that f splits, i.e. we can write $F = P_o \oplus P_1$, where $P_o = \operatorname{Ker} f$ and P_1 is mapped isomorphically onto E by f. Now the linear independence of $x_1, \ldots, x_n$ mod $\mathfrak{m}E$ shows that

$$P_o \subset \mathfrak{m}E = \mathfrak{m}P_o \subset \mathfrak{m}P_1.$$

Hence $P_o \subset \mathfrak{m}P_o$. Also, as a direct summand in a finitely generated module, P_o is finitely generated. So by Lemma 4.3, $P_o = (0)$ and f is an isomorphism, as was to be proved.

As to the second statement, it is immediate since we can complete a given

sequence $x_1, \dots, x_r$ with $\bar{x}_1, \dots, \bar{x}_r$ linearly independent over $A/\mathfrak{m}$, to a sequence $x_1, \dots, x_n$ with $\bar{x}_1, \dots, \bar{x}_n$ lineary independent over $A/\mathfrak{m}$, and then we can apply the first part of the proof. This concludes the proof of the theorem.

Let E be a module over a local ring A with maximal ideal $\mathfrak{m}$. We let $E(\mathfrak{m}) = E/\mathfrak{m}E$. If $f: E \to F$ is a homomorphism, then f induces a homomorphism

$$f_{(\mathfrak{m})}: E(\mathfrak{m}) \to F(\mathfrak{m}).$$

If f is surjective, then it follows trivially that $f_{(\mathfrak{m})}$ is surjective.

Proposition 4.5. *Let $f: E \to F$ be a homomorphism of modules, finite over a local ring A. Then:*

(i) *If $f_{(\mathfrak{m})}$ is surjective, so is f.*

(ii) *Assume f is injective. If $f_{(\mathfrak{m})}$ is surjective, then f is an isomorphism.*

(iii) *Assume that E, F are free. If $f_{(\mathfrak{m})}$ is injective (resp. an isomorphism) then f is injective (resp. an isomorphism).*

Proof. The proofs are immediate consequences of Nakayama's lemma and will be left to the reader. For instance, in the first statement, consider the exact sequence

$$E \to F \to F/\operatorname{Im} f \to 0$$

and apply Nakayama to the term on the right. In (iii), use the lifting of bases as in Theorem 4.4.

§5. FILTERED AND GRADED MODULES

Let A be a commutative ring and E a module. By a **filtration** of E one means a sequence of submodules

$$E = E_0 \supset E_1 \supset E_2 \supset \cdots \supset E_n \supset \cdots$$

Strictly speaking, this should be called a descending filtration. We don't consider any other.

Example. Let $\mathfrak{a}$ be an ideal of a ring A, and E an A-module. Let

$$E_n = \mathfrak{a}^n E.$$

Then the sequence of submodules $\{E_n\}$ is a filtration.

More generally, let $\{E_n\}$ be any filtration of a module E. We say that it is an **$\mathfrak{a}$-filtration** if $\mathfrak{a}E_n \subset E_{n+1}$ for all n. The preceding example is an $\mathfrak{a}$-filtration.

We say that an $\mathfrak{a}$-filtration is **$\mathfrak{a}$-stable**, or **stable** if we have $\mathfrak{a}E_n = E_{n+1}$ for all n sufficiently large.

Proposition 5.1. *Let $\{E_n\}$ and $\{E'_n\}$ be stable $\mathfrak{a}$-filtrations of E. Then there exists a positive integer d such that*

$$E_{n+d} \subset E'_n \quad \text{and} \quad E'_{n+d} \subset E_n$$

for all $n \geqq 0$.

Proof. It suffices to prove the proposition when $E'_n = \mathfrak{a}^n E$. Since $\mathfrak{a}E_n \subset E_{n+1}$ for all n, we have $\mathfrak{a}^n E \subset E_n$. By the stability hypothesis, there exists d such that

$$E_{n+d} = \mathfrak{a}^n E_d \subset \mathfrak{a}^n E,$$

which proves the proposition.

A ring A is called **graded** (by the natural numbers) if one can write A as a direct sum (as abelian group),

$$A = \bigoplus_{n=0}^{\infty} A_n,$$

such that for all integers $m, n \geqq 0$ we have $A_n A_m \subset A_{n+m}$. It follows in particular that A_0 is a subring, and that each component A_n is an A_0-module.

Let A be a graded ring. A module E is called a **graded module** if E can be expressed as a direct sum (as abelian group)

$$E = \bigoplus_{n=0}^{\infty} E_n,$$

such that $A_n E_m \subset E_{n+m}$. In particular, E_n is an A_0-module. Elements of E_n are then called **homogeneous of degree** n. By definition, any element of E can be written uniquely as a finite sum of homogeneous elements.

Example. Let k be a field, and let $X_0, \ldots, X_r$ be independent variables. The polynomial ring $A = k[X_0, \ldots, X_r]$ is a graded algebra, with $k = A_0$. The homogeneous elements of degree n are the polynomials generated by the monomials in $X_0, \ldots, X_r$ of degree n, that is

$$X_0^{d_0} \cdots X_r^{d_r} \quad \text{with} \quad \sum_{i=0}^{r} d_i = n.$$

An ideal I of A is called homogeneous if it is graded, as an A-module. If this is the case, then the factor ring A/I is also a graded ring.

Proposition 5.2. *Let A be a graded ring. Then A is Noetherian if and only if A_0 is Noetherian, and A is finitely generated as A_0-algebra.*

Proof. A finitely generated algebra over a Noetherian ring is Noetherian, because it is a homomorphic image of the polynomial ring in finitely many variables, and we can apply Hilbert's theorem.

Conversely, suppose that A is Noetherian. The sum

$$A^+ = \bigoplus_{n=1}^{\infty} A_n$$

is an ideal of A, whose residue class ring is A_0, which is thus a homomorphic image of A, and is therefore Noetherian. Furthermore, A^+ has a finite number of generators $x_1, \ldots, x_s$ by hypothesis. Expressing each generator as a sum of homogeneous elements, we may assume without loss of generality that these generators are homogeneous, say of degrees $d_1, \ldots, d_s$ respectively, with all $d_i > 0$. Let B be the subring of A generated over A_0 by $x_1, \ldots, x_s$. We claim that $A_n \subset B$ for all n. This is certainly true for $n = 0$. Let $n > 0$. Let x be homogeneous of degree n. Then there exist elements $a_i \in A_{n-d_i}$ such that

$$x = \sum_{i=1}^{s} a_i x_i.$$

Since $d_i > 0$ by induction, each a_i is in $A_0[x_1, \ldots, x_s] = B$, so this shows $x \in B$ also, and concludes the proof.

We shall now see two ways of constructing graded rings from filtrations.

First, let A be a ring and $\mathfrak{a}$ an ideal. We view A as a filtered ring, by the powers $\mathfrak{a}^n$. We define the **first associated graded ring** to be

$$S_{\mathfrak{a}}(A) = S = \bigoplus_{n=0}^{\infty} \mathfrak{a}^n.$$

Similarly, if E is an A-module, and E is filtered by an $\mathfrak{a}$-filtration, we define

$$E_S = \bigoplus_{n=0}^{\infty} E_n.$$

Then it is immediately verified that E_S is a graded S-module.

Observe that if A is Noetherian, and $\mathfrak{a}$ is generated by elements $x_1, \ldots, x_s$ then S is generated as an A-algebra also by $x_1, \ldots, x_s$, and is therefore also Noetherian.

Lemma 5.3. *Let A be a Noetherian ring, and E a finitely generated module, with an $\mathfrak{a}$-filtration. Then E_S is finite over S if and only if the filtration of E is $\mathfrak{a}$-stable.*

Proof. Let

$$F_n = \bigoplus_{i=0}^{n} E_i,$$

and let

$$G_n = E_0 \oplus \cdots \oplus E_n \oplus \mathfrak{a}E_n \oplus \mathfrak{a}^2E_n \oplus \mathfrak{a}^3E_n \oplus \cdots$$

Then G_n is an S-submodule of E_S, and is finite over S since F_n is finite over A. We have

$$G_n \subset G_{n+1} \quad \text{and} \quad \bigcup G_n = E_S.$$

Since S is Noetherian, we get:

$$\begin{aligned} E_S \text{ is finite over } S &\Leftrightarrow E_S = G_N \text{ for some } N \\ &\Leftrightarrow E_{N+m} = \mathfrak{a}^m E_N \text{ for all } m \geqq 0 \\ &\Leftrightarrow \text{the filtration of } E \text{ is } \mathfrak{a}\text{-stable.} \end{aligned}$$

This proves the lemma.

Theorem 5.4. (Artin-Rees). *Let A be a Noetherian ring, $\mathfrak{a}$ an ideal, E a finite A-module with a stable $\mathfrak{a}$-filtration. Let F be a submodule, and let $F_n = F \cap E_n$. Then $\{F_n\}$ is a stable $\mathfrak{a}$-filtration of F.*

Proof. We have

$$\mathfrak{a}(F \cap E_n) \subset \mathfrak{a}F \cap \mathfrak{a}E_n \subset F \cap E_{n+1},$$

so $\{F_n\}$ is an $\mathfrak{a}$-filtration of F. We can then form the associated graded S-module F_S, which is a submodule of E_S, and is finite over S since S is Noetherian. We apply Lemma 5.3 to conclude the proof.

We reformulate the Artin-Rees theorem in its original form as follows.

Corollary 5.5. *Let A be a Noetherian ring, E a finite A-module, and F a submodule. Let $\mathfrak{a}$ be an ideal. There exists an integer s such that for all integers $n \geqq s$ we have*

$$\mathfrak{a}^n E \cap F = \mathfrak{a}^{n-s}(\mathfrak{a}^s E \cap F).$$

Proof. Special case of Theorem 5.4 and the definitions.

Theorem 5.6. (Krull). *Let A be a Noetherian ring, and let $\mathfrak{a}$ be an ideal contained in every maximal ideal of A. Let E be a finite A-module. Then*

$$\bigcap_{n=1}^{\infty} \mathfrak{a}^n E = 0.$$

Proof. Let $F = \bigcap \mathfrak{a}^n E$ and apply Nakayama's lemma to conclude the proof.

Corollary 5.7. *Let $\mathfrak{o}$ be a local Noetherian ring with maximal ideal $\mathfrak{m}$. Then*

$$\bigcap_{n=1}^{\infty} \mathfrak{m}^n = 0.$$

Proof. Special case of Theorem 5.6 when $E = A$.

The second way of forming a graded ring or module is done as follows. Let A be a ring and $\mathfrak{a}$ an ideal of A. We define the **second associated graded ring**

$$\mathrm{gr}_{\mathfrak{a}}(A) = \bigoplus_{n=0}^{\infty} \mathfrak{a}^n/\mathfrak{a}^{n+1}.$$

Multiplication is defined in the obvious way. Let $a \in \mathfrak{a}^n$ and let $\bar{a}$ denote its residue class mod $\mathfrak{a}^{n+1}$. Let $b \in \mathfrak{a}^m$ and let $\bar{b}$ denote its residue class mod $\mathfrak{a}^{m+1}$. We define the product $\bar{a}\bar{b}$ to be the residue class of ab mod $\mathfrak{a}^{m+n+1}$. It is easily verified that this definition is independent of the choices of representatives and defines a multiplication on $\mathrm{gr}_{\mathfrak{a}}(A)$ which makes $\mathrm{gr}_{\mathfrak{a}}(A)$ into a graded ring.

Let E be a filtered A-module. We define

$$\mathrm{gr}(E) = \bigoplus_{n=0}^{\infty} E_n/E_{n+1}.$$

If the filtration is an $\mathfrak{a}$-filtration, then $\mathrm{gr}(E)$ is a graded $\mathrm{gr}_{\mathfrak{a}}(A)$-module.

Proposition 5.8. *Assume that A is Noetherian, and let $\mathfrak{a}$ be an ideal of A. Then $\mathrm{gr}_{\mathfrak{a}}(A)$ is Noetherian. If E is a finite A-module with a stable $\mathfrak{a}$-filtration, then $\mathrm{gr}(E)$ is a finite $\mathrm{gr}_{\mathfrak{a}}(A)$-module.*

Proof. Let $x_1, \ldots, x_s$ be generators of $\mathfrak{a}$. Let $\bar{x}_i$ be the residue class of x_i in $\mathfrak{a}/\mathfrak{a}^2$. Then

$$\mathrm{gr}_{\mathfrak{a}}(A) = (A/\mathfrak{a})[\bar{x}_1, \ldots, \bar{x}_s]$$

is Noetherian, thus proving the first assertion. For the second assertion, we have for some d,

$$E_{d+m} = \mathfrak{a}^m E_d \qquad \text{for all } m \geqq 0.$$

Hence $\mathrm{gr}(E)$ is generated by the finite direct sum

$$\mathrm{gr}(E)_0 \oplus \cdots \oplus \mathrm{gr}(E)_d.$$

But each $\mathrm{gr}(E)_n = E_n/E_{n+1}$ is finitely generated over A, and annihilated by $\mathfrak{a}$, so is a finite $A/\mathfrak{a}$-module. Hence the above finite direct sum is a finite $A/\mathfrak{a}$-module, so $\mathrm{gr}(E)$ is a finite $\mathrm{gr}_{\mathfrak{a}}(A)$-module, thus concluding the proof of the proposition.

§6. THE HILBERT POLYNOMIAL

The main point of this section is to study the lengths of certain filtered modules over local rings, and to show that they are polynomials in appropriate cases. However, we first look at graded modules, and then relate filtered modules to graded ones by using the construction at the end of the preceding section.

We start with a graded Noetherian ring together with a finite graded A-module E, so

$$A = \bigoplus_{n=0}^{\infty} A_n \quad \text{and} \quad E = \bigoplus_{n=0}^{\infty} E_n.$$

We have seen in Proposition 5.2 that A_0 is Noetherian, and that A is a finitely generated A_0-algebra. The same type of argument shows that E has a finite number of homogeneous generators, and E_n is a finite A_0-module for all $n \geqq 0$.

Let φ be an Euler-Poincaré **Z**-valued function on the class of all finite A_0-modules, as in Chapter III, §8. We define the **Poincaré series** with respect to φ to be the power series

$$P_\varphi(E, t) = \sum_{n=0}^{\infty} \varphi(E_n)t^n \in \mathbf{Z}[[t]].$$

We write $P(E, t)$ instead of $P_\varphi(E, t)$ for simplicity.

Theorem 6.1. (Hilbert-Serre). *Let s be the number of generators of A as A_0-algebra. Then $P(E, t)$ is a rational function of type*

$$P(E, t) = \frac{f(t)}{\prod_{i=1}^{s} (1 - t^{d_i})}$$

with suitable positive integers d_i, and $f(t) \in \mathbf{Z}[t]$.

Proof. Induction on s. For $s = 0$ the assertion is trivially true. Let $s \geqq 1$. Let $A = A_0[x_1, \ldots, x_s]$, deg. $x_i = d_i \geqq 1$. Multiplication by x_s on E gives rise to an exact sequence

$$0 \to K_n \to E_n \xrightarrow{x_s} E_{n+d_s} \to L_{n+d_s} \to 0.$$

Let

$$K = \bigoplus K_n \quad \text{and} \quad L = \bigoplus L_n.$$

Then K, L are finite A-modules (being submodules and factor modules of E), and are annihilated by x_s, so are in fact graded $A_0[x_1, \ldots, x_{s-1}]$-modules. By definition of an Euler-Poincaré function, we get

$$\varphi(K_n) - \varphi(E_n) + \varphi(E_{n+d_s}) - \varphi(L_{n+d_s}) = 0.$$

Multiplying by t^{n+d_s} and summing over n, we get

$$(1 - t^{d_s})P(E, t) = P(L, t) - t^{d_s}P(K, t) + g(t),$$

where $g(t)$ is a polynomial in $\mathbf{Z}[t]$. The theorem follows by induction.

Remark. In Theorem 6.1, if $A = A_0[x_1, \ldots, x_s]$ then $d_i = \deg x_i$ as shown in the proof. The next result shows what happens when all the degrees are equal to 1.

Theorem 6.2. *Assume that A is generated as an A_0-algebra by homogeneous elements of degree 1. Let d be the order of the pole of $P(E, t)$ at $t = 1$. Then for all sufficiently large n, $\varphi(E_n)$ is a polynomial in n of degree $d - 1$. (For this statement, the zero polynomial is assumed to have degree -1.)*

Proof. By Theorem 6.1, $\varphi(E_n)$ is the coefficient of t^n in the rational function

$$P(E, t) = f(t)/(1 - t)^s.$$

Cancelling powers of $1 - t$, we write $P(E, t) = h(t)/(1 - t)^d$, and $h(1) \neq 0$, with $h(t) \in \mathbf{Z}[t]$. Let

$$h(t) = \sum_{k=0}^{m} a_k t^k.$$

We have the binomial expansion

$$(1 - t)^{-d} = \sum_{k=0}^{\infty} \binom{d + k - 1}{d - 1} t^k.$$

For convenience we let $\binom{n}{-1} = 0$ for $n \geqq 0$ and $\binom{n}{-1} = 1$ for $n = -1$. We then get

$$\varphi(E_n) = \sum_{k=0}^{m} a_k \binom{d + n - k - 1}{d - 1} \quad \text{for all } n \geqq m.$$

The sum on the right-hand side is a polynomial in n with leading term

$$\left(\sum a_k\right) \frac{n^{d-1}}{(d - 1)!} \neq 0.$$

This proves the theorem.

The polynomial of Theorem 6.2 is called the **Hilbert polynomial** of the graded module E, with respect to φ.

We now put together a number of results of this chapter, and give an application of Theorem 6.2 to certain filtered modules.

Let A be a Noetherian local ring with maximal ideal $\mathfrak{m}$. Let $\mathfrak{q}$ be an $\mathfrak{m}$-primary ideal. Then $A/\mathfrak{q}$ is also Noetherian and local. Since some power of $\mathfrak{m}$ is contained in $\mathfrak{q}$, it follows that $A/\mathfrak{q}$ has only one associated prime, viewed as module over itself, namely $\mathfrak{m}/\mathfrak{q}$ itself. Similarly, if M is a finite $A/\mathfrak{q}$-module, then M has only one associated prime, and the only simple $A/\mathfrak{q}$-module is in fact an $A/\mathfrak{m}$-module which is one-dimensional. Again since some power of $\mathfrak{m}$ is contained in $\mathfrak{q}$, it follows that $A/\mathfrak{q}$ has finite length, and M also has finite length. We now use the length function as an Euler-Poincaré function in applying Theorem 6.2.

Theorem 6.3. *Let A be a Noetherian local ring with maximal ideal $\mathfrak{m}$. Let $\mathfrak{q}$ be an $\mathfrak{m}$-primary ideal, and let E be a finitely generated A-module, with a stable $\mathfrak{q}$-filtration. Then:*

(i) *E/E_n has finite length for $n \geqq 0$.*

(ii) *For all sufficiently large n, this length is a polynomial $g(n)$ of degree $\leqq s$, where s is the least number of generators of $\mathfrak{q}$.*

(iii) *The degree and leading coefficient of $g(n)$ depend only on E and $\mathfrak{q}$, but not on the chosen filtration.*

Proof. Let

$$G = \operatorname{gr}_{\mathfrak{q}}(A) = \bigoplus \mathfrak{q}^n/\mathfrak{q}^{n+1}.$$

Then $\operatorname{gr}(E) = \bigoplus E_n/E_{n+1}$ is a graded G-module, and $G_0 = A/\mathfrak{q}$. By Proposition 5.8, G is Noetherian and $\operatorname{gr}(E)$ is a finite G-module. By the remarks preceding the theorem, E/E_n has finite length, and if φ denotes the length, then

$$\varphi(E/E_n) = \sum_{j=1}^{n} \varphi(E_{j-1}/E_j).$$

If $x_1, \ldots, x_s$ generate $\mathfrak{q}$, then the images $\bar{x}_1, \ldots, \bar{x}_s$ in $\mathfrak{q}/\mathfrak{q}^2$ generate G as $A/\mathfrak{q}$-algebra, and each $\bar{x}_i$ has degree 1. By Theorem 6.2 we see that

$$\varphi(E_n/E_{n+1}) = h(n)$$

is a polynomial in n of degree $\leqq s - 1$ for sufficiently large n. Since

$$\varphi(E/E_{n+1}) - \varphi(E/E_n) = h(n),$$

it follows by Lemma 6.4 below that $\varphi(E/E_n)$ is a polynomial $g(n)$ of degree $\leqq s$ for all large n. The last statement concerning the independence of the degree

of g and its leading coefficient from the chosen filtration follows immediately from Proposition 5.1, and will be left to the reader. This concludes the proof.

From the theorem, we see that there is a polynomial $\chi_{E,\mathfrak{q}}$ such that

$$\chi_{E,\mathfrak{q}}(n) = \text{length}(E/\mathfrak{q}^n E)$$

for all sufficiently large n. If $E = A$, then $\chi_{A,\mathfrak{q}}$ is usually called the **characteristic polynomial** of $\mathfrak{q}$. In particular, we see that

$$\chi_{A,\mathfrak{q}}(n) = \text{length}(A/\mathfrak{q}^n)$$

for all sufficiently large n.

For a continuation of these topics into dimension theory, see [AtM 69] and [Mat 80].

We shall now study a particularly important special case having to do with polynomial ideals. Let k be a field, and let

$$A = k[X_0, \ldots, X_N]$$

be the polynomial ring in $N + 1$ variable. Then A is graded, the elements of degree n being the homogeneous polynomials of degree n. We let $\mathfrak{a}$ be a homogeneous ideal of A, and for an integer $n \geqq 0$ we define:

$$\begin{aligned}
\varphi(n) &= \dim_k A_n \\
\varphi(n, \mathfrak{a}) &= \dim_k \mathfrak{a}_n \\
\chi(n, \mathfrak{a}) &= \dim_k A_n/\mathfrak{a}_n = \dim_k A_n - \dim_k \mathfrak{a}_n = \varphi(n) - \varphi(n, \mathfrak{a}).
\end{aligned}$$

As earlier in this section, A_n denotes the k-space of homogeneous elements of degree n in A, and similarly for $\mathfrak{a}_n$. Then we have

$$\varphi(n) = \binom{N + n}{N}.$$

We shall consider the **binomial polynomial**

$$\binom{T}{d} = \frac{T(T-1)\cdots(T-d+1)}{d!} = \frac{T^d}{d!} + \text{lower terms}. \tag{1}$$

If f is a function, we define the **difference function** Δf by

$$\Delta f(T) = f(T+1) - f(T).$$

Then one verifies directly that

$$\Delta\binom{T}{d} = \binom{T}{d-1}. \tag{2}$$

Lemma 6.4. *Let $P \in \mathbf{Q}[T]$ be a polynomial of degree d with rational coefficients.*

(a) *If $P(n) \in \mathbf{Z}$ for all sufficiently large integers n, then there exist integers $c_0, \ldots, c_d$ such that*

$$P(T) = c_0\binom{T}{d} + c_1\binom{T}{d-1} + \ldots + c_d.$$

In particular, $P(n) \in \mathbf{Z}$ for all integers n.

(b) *If $f\colon \mathbf{Z} \to \mathbf{Z}$ is any function, and if there exists a polynomial $Q(T) \in \mathbf{Q}[T]$ such that $Q(\mathbf{Z}) \subset \mathbf{Z}$ and $\Delta f(n) = Q(n)$ for all n sufficiently large, then there exists a polynomial P as in (a) such that $f(n) = P(n)$ for all n sufficiently large.*

Proof. We prove (a) by induction. If the degree of P is 0, then the assertion is obvious. Suppose $\deg P \geqq 1$. By (1) there exist rational numbers $c_0, \ldots, c_d$ such that $P(T)$ has the expression given in (a). But ΔP has degree strictly smaller than $\deg P$. Using (2) and induction, we conclude that $c_0, \ldots, c_{d-1}$ must be integers. Finally c_d is an integer because $P(n) \in \mathbf{Z}$ for n sufficiently large. This proves (a).

As for (b), using (a), we can write

$$Q(T) = c_0\binom{T}{d-1} + \ldots + c_{d-1}$$

with integers $c_0, \ldots, c_{d-1}$. Let P_1 be the "integral" of Q, that is

$$P_1(T) = c_0\binom{T}{d} + \ldots + c_{d-1}\binom{T}{1}, \quad \text{so} \quad \Delta P_1 = Q.$$

Then $\Delta(f - P_1)(n) = 0$ for all n sufficiently large. Hence $(f - P_1)(n)$ is equal to a constant c_d for all n sufficiently large, so we let $P = P_1 + c_d$ to conclude the proof.

Proposition 6.5. *Let $\mathfrak{a}, \mathfrak{b}$ be homogeneous ideals in A. Then*

$$\varphi(n, \mathfrak{a} + \mathfrak{b}) = \varphi(n, \mathfrak{a}) + \varphi(n, \mathfrak{b}) - \varphi(n, \mathfrak{a} \cap \mathfrak{b})$$
$$\chi(n, \mathfrak{a} + \mathfrak{b}) = \chi(n, \mathfrak{a}) + \chi(n, \mathfrak{b}) - \chi(n, \mathfrak{a} \cap \mathfrak{b}).$$

Proof. The first is immediate, and the second follows from the definition of χ.

Theorem 6.6. *Let F be a homogeneous polynomial of degree d. Assume that F is not a divisor of zero mod $\mathfrak{a}$, that is: if $G \in A$, $FG \in \mathfrak{a}$, then $G \in \mathfrak{a}$. Then*

$$\chi(n, \quad + (F)) = \chi(n, \mathfrak{a}) - \chi(n - d, \mathfrak{a}).$$

Proof. First observe that trivially

$$\varphi(n, (F)) = \varphi(n - d),$$

because the degree of a product is the sum of the degrees. Next, using the hypothesis that F is not divisor of 0 mod $\mathfrak{a}$, we conclude immediately

$$\varphi(n, \mathfrak{a} \cap (F)) = \varphi(n - d, \mathfrak{a}).$$

Finally, by Proposition 6.5 (the formula for χ), we obtain:

$$\begin{aligned} \chi(n, \mathfrak{a} + (F)) &= \chi(n, \mathfrak{a}) + \chi(n, (F)) - \chi(n, \mathfrak{a} \cap (F)) \\ &= \chi(n, \mathfrak{a}) + \varphi(n) - \varphi(n, (F)) - \varphi(n) + \varphi(n, \mathfrak{a} \cap (F)) \\ &= \chi(n, \mathfrak{a}) - \varphi(n - d) + \varphi(n - d, \mathfrak{a}) \\ &= \chi(n, \mathfrak{a}) - \chi(n - d, \mathfrak{a}) \end{aligned}$$

thus proving the theorem.

We denote by $\mathfrak{m}$ the maximal ideal $\mathfrak{m} = (X_0, \ldots, X_N)$ in A. We call $\mathfrak{m}$ the **irrelevant prime ideal**. An ideal is called **irrelevant** if some positive power of $\mathfrak{m}$ is contained in the ideal. In particular, a primary ideal $\mathfrak{q}$ is irrelevant if and only if $\mathfrak{m}$ belongs to $\mathfrak{q}$. Note that by the Hilbert nullstellensatz, the condition that some power of $\mathfrak{m}$ is contained in $\mathfrak{a}$ is equivalent with the condition that the only zero of $\mathfrak{a}$ (in some algebraically closed field containing k) is the trivial zero.

Proposition 6.7. *Let $\mathfrak{a}$ be a homogeneous ideal.*

(a) *If $\mathfrak{a}$ is irrelevant, then $\chi(n, \mathfrak{a}) = 0$ for n sufficiently large.*

(b) *In general, there is an expression $\mathfrak{a} = \mathfrak{q}_1 \cap \ldots \cap \mathfrak{q}_s$ as a reduced primary decomposition such that all $\mathfrak{q}_i$ are homogeneous.*

(c) *If an irrelevant primary ideal occurs in the decomposition, let $\mathfrak{b}$ be the intersection of all other primary ideals. Then*

$$\chi(n, \mathfrak{a}) = \chi(n, \mathfrak{b})$$

for all n sufficiently large.

Proof. For (a), by assumption we have $A_n = \mathfrak{a}_n$ for n sufficiently large, so the assertion (a) is obvious. We leave (b) as an exercise. As to (c), say $\mathfrak{q}_s$ is irrelevant, and let $\mathfrak{b} = \mathfrak{q}_1 \cap \ldots \cap \mathfrak{q}_{s-1}$. By Proposition 6.5, we have

$$\chi(n, \mathfrak{b} + \mathfrak{q}_s) = \chi(n, \mathfrak{b}) + \chi(n, \mathfrak{q}_s) - \chi(n, \mathfrak{a}).$$

But $\mathfrak{b} + \mathfrak{q}_s$ is irrelevant, so (c) follows from (a), thus concluding the proof.

We now want to see that for any homogeneous ideal $\mathfrak{a}$ the function f such that

$$f(n) = \chi(n, \mathfrak{a})$$

satisfies the conditions of Lemma 6.4(b). First, we observe that if we change the ground field from k to an algebraically closed field K containing k, and we let $A_K = K[X_0, \ldots, X_N]$, $\mathfrak{a}_K = K\mathfrak{a}$, then

$$\dim_k A_n = \dim_K A_{K,n} \qquad \text{and} \qquad \dim_k \mathfrak{a}_n = \dim_K \mathfrak{a}_{K,n}.$$

Hence we can assume that k is algebraically closed.

Second, we shall need a geometric notion, that of dimension. Let V be a variety over k, say affine, with generic point $(x) = (x_1, \ldots, x_N)$. We define its **dimension** to be the transcendence degree of $k(x)$ over k. For a projective variety, defined by a homogeneous prime ideal $\mathfrak{p}$, we define its dimension to be the dimension of the homogeneous variety defined by $\mathfrak{p}$ minus 1.

We now need the following lemma.

Lemma 6.8. *Let V, W be varieties over a field k.*

If $V \supset W$ and $\dim V = \dim W$, *then $V = W$.*

Proof. Say V, W are in affine space $\mathbf{A}^N$. Let $\mathfrak{p}_V$ and $\mathfrak{p}_W$ be the respective prime ideals of V and W in $k[X]$. Then we have a canonical homomorphism

$$k[X]/\mathfrak{p}_V \approx k[x] \to k[y] \approx k[X]/\mathfrak{p}_W$$

from the affine coordinate ring of V onto the affine coordinate ring of W. If the transcendence degree of $k(x)$ is the same as that of $k(y)$, and say $y_1, \ldots, y_r$ form a transcendence basis of $k(y)$ over k, then $x_1, \ldots, x_r$ is a transcendence basis of $k(x)$ over k, the homomorphism $k[x] \to k[y]$ induces an isomorphism

$$k[x_1, \ldots, x_r] \xrightarrow{\approx} k[y_1, \ldots, y_r],$$

and hence an isomorphism on the finite extension $k[x]$ to $k[y]$, as desired.

Theorem 6.9. *Let $\mathfrak{a}$ be a homogeneous ideal in A. Let r be the maximum dimension of the irreducible components of the algebraic space in projective space defined by $\mathfrak{a}$. Then there exists a polynomial $P \in \mathbf{Q}[T]$ of degree $\leqq r$, such that $P(\mathbf{Z}) \subset \mathbf{Z}$, and such that*

$$P(n) = \chi(n, \mathfrak{a})$$

for all n sufficiently large.

Proof. By Proposition 6.7(c), we may assume that no primary component in the primary decomposition of $\mathfrak{a}$ is irrelevant. Let Z be the algebraic space of zeros of $\mathfrak{a}$ in projective space. We may assume k algebraically closed as noted previously. Then there exists a homogeneous polynomial $L \in k[X]$ of degree 1 (a linear form) which does not lie in any of the prime ideals belonging to the primary ideals in the given decomposition. In particular, L is not a divisor of zero mod $\mathfrak{a}$. Then the components of the algebraic space of zeros of $\mathfrak{a} + (L)$ must have dimension $\leqq r - 1$. By induction and Theorem 6.6, we conclude that the difference

$$\chi(n, \mathfrak{a}) - \chi(n - 1, \mathfrak{a})$$

satisfies the conditions of Lemma 6.4(b), which concludes the proof.

The polynomial in Theorem 6.9 is called the **Hilbert polynomial** of the ideal $\mathfrak{a}$.

Remark. The above results give an introduction for Hartshorne's [Ha 77], Chapter I, especially §7. If Z is not empty, and if we write

$$\chi(n, \mathfrak{a}) = c\frac{n^r}{r!} + \text{lower terms},$$

then $c > 0$ and c can be interpreted as the **degree** of Z, or in geometric terms, the number of points of intersection of Z with a sufficiently general linear variety of complementary dimension (counting the points with certain multiplicities). For explanations and details, see [Ha 77], Chapter I, Proposition 7.6 and Theorem 7.7; van der Waerden [vdW 29] which does the same thing for multihomogeneous polynomial ideals; [La 58], referred to at the end of Chapter VIII, §2; and the papers [MaW 85], [Ph 86], making the link with van der Waerden some six decades before.

Bibliography

[AtM 69] M. ATIYAH and I. MACDONALD, *Introduction to commutative algebra*, Addison-Wesley, 1969

[Ha 77] R. HARTSHORNE, *Algebraic Geometry*, Springer Verlag, 1977

[MaW 85] D. MASSER and G. WÜSTHOLZ, Zero estimates on group varieties II, *Invent. Math.* **80** (1985), pp. 233–267

[Mat 80] H. MATSUMURA, *Commutative algebra*, Second Edition, Benjamin-Cummings, 1980

[Ph 86] P. PHILIPPON, Lemmes de zéros dans les groupes algébriques commutatifs, *Bull. Soc. Math. France* **114** (1986), pp. 355–383

[vdW 29] B. L. VAN DER WAERDEN, On Hilbert's function, series of composition of ideals and a generalization of the theorem of Bezout, *Proc. R. Soc. Amsterdam* **31** (1929), pp. 749–770

§7. INDECOMPOSABLE MODULES

Let A be a ring, not necessarily commutative, and E an A-module. We say that E is **Artinian** if E satisfies the descending chain condition on submodules, that is a sequence

$$E_1 \supset E_2 \supset E_3 \cdots$$

must stabilize: there exists an integer N such that if $n \geqq N$ then $E_n = E_{n+1}$.

Example 1. If k is a field, A is a k-algebra, and E is a finite-dimensional vector space over k which is also an A-module, then E is Artinian as well as Noetherian.

Example 2. Let A be a commutative Noetherian local ring with maximal ideal $\mathfrak{m}$, and let $\mathfrak{q}$ be an $\mathfrak{m}$-primary ideal. Then for every positive integer n, $A/\mathfrak{q}^n$ is Artinian. Indeed, $A/\mathfrak{q}^n$ has a Jordan-Hölder filtration in which each factor is a finite dimensional vector space over the field $A/\mathfrak{m}$, and is a module of finite length. See Proposition 7.2.

Conversely, suppose that A is a local ring which is both Noetherian and Artinian. Let $\mathfrak{m}$ be the maximal ideal. Then there exists some positive integer n such that $\mathfrak{m}^n = 0$. Indeed, the descending sequence $\mathfrak{m}^n$ stabilizes, and Nakayama's lemma implies our assertion. It then also follows that every primary ideal is nilpotent.

As with Noetherian rings and modules, it is easy to verify the following statements:

Proposition 7.1. *Let A be a ring, and let*

$$0 \to E' \to E \to E'' \to 0$$

be an exact sequence of A-modules. Then E is Artinian if and only if E' and E'' are Artinian.

We leave the proof to the reader. The proof is the same as in the Noetherian case, reversing the inclusion relations between modules.

Proposition 7.2. *A module E has a finite simple filtration if and only if E is both Noetherian and Artinian.*

Proof. A simple module is generated by one element, and so is Noetherian. Since it contains no proper submodule $\neq 0$, it is also Artinian. Proposition 7.2 is then immediate from Proposition 7.1.

A module E is called **decomposable** if E can be written as a direct sum

$$E = E_1 \oplus E_2$$

with $E_1 \neq E$ and $E_2 \neq E$. Otherwise, E is called **indecomposable**. If E is decomposable as above, let e_1 be the projection on the first factor, and $e_2 = 1 - e_1$ the projection on the second factor. Then e_1, e_2 are idempotents such that

$$e_1 \neq 1, \qquad e_2 \neq 1, \qquad e_1 + e_2 = 1 \quad \text{and} \quad e_1 e_2 = e_2 e_1 = 0.$$

Conversely, if such idempotents exist in $\text{End}(E)$ for some module E, then E is decomposable, and e_i is the projection on the submodule $e_i E$.

Let $u: E \to E$ be an endomorphism of some module E. We can form the descending sequence

$$\text{Im } u \supset \text{Im } u^2 \supset \text{Im } u^3 \supset \cdots$$

If E is Artinian, this sequence stabilizes, and we have

$$\text{Im } u^n = \text{Im } u^{n+1} \qquad \text{for all sufficiently large } n.$$

We call this submodule $u^\infty(E)$, or $\text{Im } u^\infty$.

Similarly, we have an ascending sequence

$$\text{Ker } u \subset \text{Ker } u^2 \subset \text{Ker } u^3 \subset \cdots$$

which stabilizes if E is Noetherian, and in this case we write

$$\text{Ker } u^\infty = \text{Ker } u^n \qquad \text{for } n \text{ sufficiently large.}$$

Proposition 7.3. (Fitting's Lemma). *Assume that E is Noetherian and Artinian. Let $u \in \text{End}(E)$. Then E has a direct sum decomposition*

$$E = \text{Im } u^\infty \oplus \text{Ker } u^\infty.$$

Furthermore, the restriction of u to $\text{Im } u^\infty$ is an automorphism, and the restriction of u to $\text{Ker } u^\infty$ is nilpotent.

Proof. Choose n such that $\text{Im } u^\infty = \text{Im } u^n$ and $\text{Ker } u^\infty = \text{Ker } u^n$. We have

$$\text{Im } u^\infty \cap \text{Ker } u^\infty = \{0\},$$

for if x lies in the intersection, then $x = u^n(y)$ for some $y \in E$, and then $0 = u^n(x) = u^{2n}(y)$. So $y \in \text{Ker } u^{2n} = \text{Ker } u^n$, whence $x = u^n(y) = 0$.

Secondly, let $x \in E$. Then for some $y \in u^n(E)$ we have

$$u^n(x) = u^n(y).$$

Then we can write

$$x = x - u^n(y) + u^n(y),$$

which shows that $E = \operatorname{Im} u^\infty + \operatorname{Ker} u^\infty$. Combined with the first step of the proof, this shows that E is a direct sum as stated.

The final assertion is immediate, since the restriction of u to $\operatorname{Im} u^\infty$ is surjective, and its kernel is 0 by the first part of the proof. The restriction of u to $\operatorname{Ker} u^\infty$ is nilpotent because $\operatorname{Ker} u^\infty = \operatorname{Ker} u^n$. This concludes the proof of the proposition.

We now generalize the notion of a local ring to a non-commutative ring. A ring A is called **local** if the set of non-units is a two-sided ideal.

Proposition 7.4. *Let E be an indecomposable module over the ring A. Assume E Noetherian and Artinian. Any endomorphism of E is either nilpotent or an automorphism. Furthermore* $\operatorname{End}(E)$ *is local.*

Proof. By Fitting's lemma, we know that for any endomorphism u, we have $E = \operatorname{Im} u^\infty$ or $E = \operatorname{Ker} u^\infty$. So we have to prove that $\operatorname{End}(E)$ is local. Let u be an endomorphism which is not a unit, so u is nilpotent. For any endomorphism v it follows that uv and vu are not surjective or injective respectively, so are not automorphisms. Let u_1, u_2 be endomorphisms which are not units. We have to show $u_1 + u_2$ is not a unit. If it is a unit in $\operatorname{End}(E)$, let $v_i = u_i(u_1 + u_2)^{-1}$. Then $v_1 + v_2 = 1$. Furthermore, $v_1 = 1 - v_2$ is invertible by the geometric series since v_2 is nilpotent. But v_1 is not a unit by the first part of the proof, contradiction. This concludes the proof.

Theorem 7.5. (Krull-Remak-Schmidt). *Let $E \neq 0$ be a module which is both Noetherian and Artinian. Then E is a finite direct sum of indecomposable modules. Up to a permutation, the indecomposable components in such a direct sum are uniquely determined up to isomorphism.*

Proof. The existence of a direct sum decomposition into indecomposable modules follows from the Artinian condition. If first $E = E_1 \oplus E_2$, then either E_1, E_2 are indecomposable, and we are done; or, say, E_1 is decomposable. Repeating the argument, we see that we cannot continue this decomposition indefinitely without contradicting the Artinian assumption.

There remains to prove uniqueness. Suppose

$$E = E_1 \oplus \cdots \oplus E_r = F_1 \oplus \cdots \oplus F_s$$

where E_i, F_j are indecomposable. We have to show that $r = s$ and after some permutation, $E_i \approx F_i$. Let e_i be the projection of E on E_i, and let u_j be the projection of E on F_j, relative to the above direct sum decompositions. Let:

$$v_j = e_1 u_j \quad \text{and} \quad w_j = u_j e_1.$$

Then $\sum u_j = \text{id}_E$ implies that

$$\sum_{j=1}^{s} v_j w_j | E_1 = \text{id}_{E_1}.$$

By Proposition 7.4, $\text{End}(E_1)$ is local, and therefore some $v_j w_j$ is an automorphism of E_1. After renumbering, we may assume that $v_1 w_1$ is an automorphism of E_1. We claim that v_1 and w_1 induce isomorphisms between E_1 and F_1, This follows from a lemma.

Lemma 7.6. *Let M, N be modules, and assume N indecomposable. Let $u: M \to N$ and $v: N \to M$ be such that vu is an automorphism. Then u, v are isomorphisms.*

Proof. Let $e = u(vu)^{-1}v$. Then $e^2 = e$ is an idempotent, lying in $\text{End}(N)$, and therefore equal to 0 or 1 since N is assumed indecomposable. But $e \neq 0$ because $\text{id}_M \neq 0$ and

$$0 \neq \text{id}_M = \text{id}_M^2 = (vu)^{-1}vu(vu)^{-1}vu.$$

So $e = \text{id}_N$. Then u is injective because vu is an automorphism; v is injective because $e = \text{id}_N$ is injective; u is surjective because $e = \text{id}_N$; and v is surjective because vu is an automorphism. This concludes the proof of the lemma.

Returning to the theorem, we now see that

$$E = F_1 \oplus (E_2 \oplus \cdots \oplus E_r).$$

Indeed, e_1 induces an isomorphism from F_1 to E_1, and since the kernel of e_1 is $E_2 \oplus \cdots \oplus E_r$ it follows that

$$F_1 \cap (E_2 \oplus \cdots \oplus E_r) = 0.$$

But also, $F_1 \equiv E_1 \pmod{E_2 \oplus \cdots \oplus E_r}$, so E is the sum of F_1 and $E_2 \oplus \cdots \oplus E_r$, whence E is the direct sum, as claimed. But then

$$E/F_1 \approx F_2 \oplus \cdots \oplus F_s \approx E_2 \oplus \cdots \oplus E_r.$$

The proof is then completed by induction.

We apply the preceding results to a commutative ring A. We note that an idempotent in A as a ring is the same thing as an idempotent as an element of $\text{End}(A)$, viewing A as module over itself. Furthermore $\text{End}(A) \approx A$. Therefore, we find the special cases:

Theorem 7.7. *Let A be a Noetherian and Artinian commutative ring.*

(i) *If A is indecomposable as a ring, then A is local.*

(ii) *In general, A is a direct product of local rings, which are Artinian and Noetherian.*

Another way of deriving this theorem will be given in the exercises.

EXERCISES

1. Let A be a commutative ring. Let M be a module, and N a submodule. Let $N = Q_1 \cap \cdots \cap Q_r$ be a primary decomposition of N. Let $\bar{Q}_i = Q_i/N$. Show that $0 = \bar{Q}_1 \cap \cdots \cap \bar{Q}_r$ is a primary decomposition of 0 in M/N. State and prove the converse.
2. Let $\mathfrak{p}$ be a prime ideal, and $\mathfrak{a}, \mathfrak{b}$ ideals of A. If $\mathfrak{a}\mathfrak{b} \subset \mathfrak{p}$, show that $\mathfrak{a} \subset \mathfrak{p}$ or $\mathfrak{b} \subset \mathfrak{p}$.
3. Let $\mathfrak{q}$ be a primary ideal. Let $\mathfrak{a}, \mathfrak{b}$ be ideals, and assume $\mathfrak{a}\mathfrak{b} \subset \mathfrak{q}$. Assume that $\mathfrak{b}$ is finitely generated. Show that $\mathfrak{a} \subset \mathfrak{q}$ or there exists some positive integer n such that $\mathfrak{b}^n \subset \mathfrak{q}$.
4. Let A be Noetherian, and let $\mathfrak{q}$ be a $\mathfrak{p}$-primary ideal. Show that there exists some $n \geqq 1$ such that $\mathfrak{p}^n \subset \mathfrak{q}$.
5. Let A be an arbitrary commutative ring and let S be a multiplicative subset. Let $\mathfrak{p}$ be a prime ideal and let $\mathfrak{q}$ be a $\mathfrak{p}$-primary ideal. Then $\mathfrak{p}$ intersects S if and only if $\mathfrak{q}$ intersects S. Furthermore, if $\mathfrak{q}$ does not intersect S, then $S^{-1}\mathfrak{q}$ is $S^{-1}\mathfrak{p}$-primary in $S^{-1}A$.
6. If $\mathfrak{a}$ is an ideal of A, let $\mathfrak{a}_S = S^{-1}\mathfrak{a}$. If $\varphi_S : A \to S^{-1}A$ is the canonical map, abbreviate $\varphi_S^{-1}(\mathfrak{a}_S)$ by $\mathfrak{a}_S \cap A$, even though φ_S is not injective. Show that there is a bijection between the prime ideals of A which do not intersect S and the prime ideals of $S^{-1}A$, given by

$$\mathfrak{p} \mapsto \mathfrak{p}_S \quad \text{and} \quad \mathfrak{p}_S \mapsto \mathfrak{p}_S \cap A = \mathfrak{p}.$$

   Prove a similar statement for primary ideals instead of prime ideals.
7. Let $\mathfrak{a} = \mathfrak{q}_1 \cap \cdots \cap \mathfrak{q}_r$ be a reduced primary decomposition of an ideal. Assume that $\mathfrak{q}_1, \ldots, \mathfrak{q}_i$ do not intersect S, but that $\mathfrak{q}_j$ intersects S for $j > i$. Show that

$$\mathfrak{a}_S = \mathfrak{q}_{1S} \cap \cdots \cap \mathfrak{q}_{iS}$$

   is a reduced primary decomposition of $\mathfrak{a}_S$.
8. Let A be a local ring. Show that any idempotent $\neq 0$ in A is necessarily the unit element. (An **idempotent** is an element $e \in A$ such that $e^2 = e$.)
9. Let A be an Artinian commutative ring. Prove:
   (a) All prime ideals are maximal. [*Hint*: Given a prime ideal $\mathfrak{p}$, let $x \in A$, $x(\mathfrak{p}) = 0$. Consider the descending chain $(x) \supset (x^2) \supset (x^3) \supset \cdots$.]

(b) There is only a finite number of prime, or maximal, ideals. [*Hint*: Among all finite intersections of maximal ideals, pick a minimal one.]
(c) The ideal N of nilpotent elements in A is nilpotent, that is there exists a positive integer k such that $N^k = (0)$. [*Hint*: Let k be such that $N^k = N^{k+1}$. Let $\mathfrak{a} = N^k$. Let $\mathfrak{b}$ be a minimal ideal $\neq 0$ such that $\mathfrak{b}\mathfrak{a} \neq 0$. Then $\mathfrak{b}$ is principal and $\mathfrak{b}\mathfrak{a} = \mathfrak{b}$.]
(d) A is Noetherian.
(e) There exists an integer r such that

$$A = \prod A/\mathfrak{m}^r$$

where the product is taken over all maximal ideals.
(f) We have

$$A = \prod A_{\mathfrak{p}},$$

where again the product is taken over all prime ideals $\mathfrak{p}$.

10. Let A, B be local rings with maximal ideals $\mathfrak{m}_A$, $\mathfrak{m}_B$, respectively. Let $f: A \to B$ be a homomorphism. We say that f is **local** if $f^{-1}(\mathfrak{m}_B) = \mathfrak{m}_A$. Suppose this is the case. Assume A, B Noetherian, and assume that:

    1. $A/\mathfrak{m}_A \to B/\mathfrak{m}_B$ is an isomorphism,

    2. $\mathfrak{m}_A \to \mathfrak{m}_B/\mathfrak{m}_B^2$ is surjective;

    3. B is a finite A-module, via f.

    Prove that f is surjective. [*Hint*: Apply Nakayama twice.]

    For an ideal $\mathfrak{a}$, recall from Chapter IX, §5 that $\mathfrak{X}(\mathfrak{a})$ is the set of primes containing $\mathfrak{a}$.

11. Let A be a commutative ring and M an A-module. Define the **support** of M by

$$\text{supp}(M) = \{\mathfrak{p} \in \text{spec}(A) : M_{\mathfrak{p}} \neq 0\}.$$

If M is finite over A, show that $\text{supp}(M) = \mathfrak{X}(\text{ann}(M))$, where $\text{ann}(M)$ is the annihilator of M in A, that is the set of elements $a \in A$ such that $aM = 0$.

12. Let A be a Noetherian ring and M a finite A-module. Let I be an ideal of A such that $\text{supp}(M) \subset \mathfrak{X}(I)$. Then $I^nM = 0$ for some $n > 0$.

13. Let A be any commutative ring, and M, N modules over A. If M is finitely presented, and S is a multiplicative subset of A, show that

$$S^{-1}\,\text{Hom}_A(M, N) \approx \text{Hom}_{S^{-1}A}(S^{-1}M, S^{-1}N).$$

This is usually applied when A is Noetherian and M finitely generated, in which case M is also finitely presented since the module of relations is a submodule of a finitely generated free module.

14. (a) Prove Proposition 6.7(b).
    (b) Prove that the degree of the polynomial P in Theorem 6.9 is exactly r.

Locally constant dimensions

15. Let A be a Noetherian local ring. Let E be a finite A-module. Assume that A has no nilpotent elements. For each prime ideal $\mathfrak{p}$ of A, let $k(\mathfrak{p})$ be the residue class field. If $\dim_{k(\mathfrak{p})} E_{\mathfrak{p}}/\mathfrak{p}E_{\mathfrak{p}}$ is constant for all $\mathfrak{p}$, show that E is free. [*Hint*: Let $x_1, \ldots, x_r \in A$ be

such that the residue classes mod the maximal ideal form a basis for $E/\mathfrak{m}E$ over $k(\mathfrak{m})$. We get a surjective homomorphism

$$A^r \to E \to 0.$$

Let J be the kernel. Show that $J_\mathfrak{p} \subset \mathfrak{m}_\mathfrak{p} A_\mathfrak{p}^r$ for all $\mathfrak{p}$ so $J \subset \mathfrak{p}$ for all $\mathfrak{p}$ and $J = 0$.]

16. Let A be a Noetherian local ring without nilpotent elements. Let $f: E \to F$ be a homomorphism of A-modules, and suppose E, F are finite free. For each prime $\mathfrak{p}$ of A let

$$f_{(\mathfrak{p})}: E_\mathfrak{p}/\mathfrak{p}E_\mathfrak{p} \to F_\mathfrak{p}/\mathfrak{p}F_\mathfrak{p}$$

be the corresponding $k(\mathfrak{p})$-homomorphism, where $k(\mathfrak{p}) = A_\mathfrak{p}/\mathfrak{p}A_\mathfrak{p}$ is the residue class field at $\mathfrak{p}$. Assume that

$$\dim_{k(\mathfrak{p})} \operatorname{Im} f_{(\mathfrak{p})}$$

is constant.

(a) Prove that $F/\operatorname{Im} f$ and $\operatorname{Im} f$ are free, and that there is an isomorphism

$$F \approx \operatorname{Im} f \oplus (F/\operatorname{Im} f).$$

[*Hint*: Use Exercise 15.]

(b) Prove that $\operatorname{Ker} f$ is free and $E \approx (\operatorname{Ker} f) \oplus (\operatorname{Im} f)$. [*Hint*: Use that finite projective is free.]

The next exercises depend on the notion of a complex, which we have not yet formally defined. A **(finite) complex** E is a sequence of homomorphisms of modules

$$0 \to E^0 \xrightarrow{d^0} E^1 \xrightarrow{d^1} \cdots \xrightarrow{d^n} E^n \to 0$$

and homorphisms $d^i: E^i \to E^{i+1}$ such that $d^{i+1} \circ d^i = 0$ for all i. Thus $\operatorname{Im}(d^i) \subset \operatorname{Ker}(d^{i+1})$. The **homology** H^i of the complex is defined to be

$$H^i = \operatorname{Ker}(d^{i+1})/\operatorname{Im}(d^i).$$

By definition, $H^0 = E^0$ and $H^n = E^n/\operatorname{Im}(d^n)$. You may want to look at the first section of Chapter XX, because all we use here is the basic notion, and the following property, which you can easily prove. Let E, F be two complexes. By a **homomorphism** $f: E \to F$ we mean a sequence of homomorphisms

$$f_i: E^i \to F^i$$

making the diagram commutative for all i:

$$\begin{array}{ccc} E^i & \xrightarrow{d_E^i} & E^{i+1} \\ f_i \uparrow & & \uparrow f_{i+1} \\ F^i & \xrightarrow[d_F^i]{} & F^{i+1} \end{array}$$

Show that such a homomorphism f induces a homomorphism $H(f): H(E) \to H(F)$ on the homology; that is, for each i we have an induced homomorphism

$$H^i(f): H^i(E) \to H^i(F).$$

The following exercises are inspired from applications to algebraic geometry, as for instance in Hartshorne, *Algebraic Geometry*, Chapter III, Theorem 12.8. See also Chapter XXI, §1 to see how one can construct complexes such as those considered in the next exercises in order to compute the homology with respect to less tractable complexes.

Reduction of a complex mod $\mathfrak{p}$

17. Let $0 \to K^0 \to K^1 \to \cdots \to K^n \to 0$ be a complex of finite free modules over a local Noetherian ring A without nilpotent elements. For each prime $\mathfrak{p}$ of A and module E, let $E(\mathfrak{p}) = E_{\mathfrak{p}}/\mathfrak{p}E_{\mathfrak{p}}$, and similarly let $K(\mathfrak{p})$ be the complex localized and reduced mod $\mathfrak{p}$. For a given integer i, assume that

$$\dim_{k(\mathfrak{p})} H^i(K(\mathfrak{p}))$$

is constant, where H^i is the i-th homology of the reduced complex. Show that $H^i(K)$ is free and that we have a natural isomorphism

$$H^i(K)(\mathfrak{p}) \xrightarrow{\approx} H^i(K(\mathfrak{p})).$$

[*Hint*: First write $d^i_{(\mathfrak{p})}$ for the map induced by d^i on $K^i(\mathfrak{p})$. Write

$$\dim_{k(\mathfrak{p})} \operatorname{Ker} d^i_{(\mathfrak{p})} = \dim_{k(\mathfrak{p})} K^i(\mathfrak{p}) - \dim_{k(\mathfrak{p})} \operatorname{Im} d^i_{(\mathfrak{p})}.$$

Then show that the dimensions $\dim_{k(\mathfrak{p})} \operatorname{Im} d^i_{(\mathfrak{p})}$ and $\dim_{k(\mathfrak{p})} \operatorname{Im} d^{i-1}_{(\mathfrak{p})}$ must be constant. Then apply Exercise 12.]

Comparison of homology at the special point

18. Let A be a Noetherian local ring. Let K be a finite complex, as follows:

$$0 \to K^0 \to \cdots \to K^n \to 0,$$

such that K^i is *finite free* for all i. For some index i assume that

$$H^i(K)(\mathfrak{m}) \to H^i(K(\mathfrak{m}))$$

is surjective. Prove:

(a) This map is an isomorphism.

(b) The following exact sequences split:

$$0 \to \operatorname{Ker} d^i \to K^i \to \operatorname{Im} d^i \to 0$$

$$0 \to \operatorname{Im} d^i \to K^{i+1}$$

(c) Every term in these sequences is free.

19. Let A be a Noetherian local ring. Let K be a complex as in the previous exercise. For some i assume that

$$H^i(K)(\mathfrak{m}) \to H^i(K(\mathfrak{m}))$$

is surjective (or equivalently is an isomorphism by the previous exercise). Prove that

the following conditions are equivalent:

(a) $H^{i-1}(K)(\mathfrak{m}) \to H^{i-1}(K(\mathfrak{m}))$ is surjective.

(b) $H^{i-1}(K)(\mathfrak{m}) \to H^{i-1}(K(\mathfrak{m}))$ is an isomorphism.

(c) $H^i(K)$ is free.

[*Hint*: Lift bases until you are blue in the face.]

(d) If these conditions hold, then each one of the two inclusions

$$\operatorname{Im} d^{i-1} \subset \operatorname{Ker} d^i \subset K^i$$

splits, and each one of these modules is free. Reducing mod $\mathfrak{m}$ yields the corresponding inclusions

$$\operatorname{Im} d^{i-1}_{(\mathfrak{m})} \subset \operatorname{Ker} d^i_{(\mathfrak{m})} \subset K^i(\mathfrak{m}),$$

and induce the isomorphism on cohomology as stated in (b). [*Hint*: Apply the preceding exercise.]

CHAPTER XI

Real Fields

§1. ORDERED FIELDS

Let K be a field. An **ordering** of K is a subset P of K having the following properties:

ORD 1. Given $x \in K$, we have either $x \in P$, or $x = 0$, or $-x \in P$, and these three possibilities are mutually exclusive. In other words, K is the disjoint union of P, $\{0\}$, and $-P$.

ORD 2. If $x, y \in P$, then $x + y$ and $xy \in P$.

We shall also say that K is **ordered by** P, and we call P the set of **positive elements**.

Let us assume that K is ordered by P. Since $1 \neq 0$ and $1 = 1^2 = (-1)^2$ we see that $1 \in P$. By **ORD 2**, it follows that $1 + \cdots + 1 \in P$, whence K has characteristic 0. If $x \in P$, and $x \neq 0$, then $xx^{-1} = 1 \in P$ implies that $x^{-1} \in P$.

Let $x, y \in K$. We define $x < y$ (or $y > x$) to mean that $y - x \in P$. If $x < 0$ we say that x is **negative**. This means that $-x$ is positive. One verifies trivially the usual relations for inequalities, for instance:

$$\begin{aligned} x < y \quad &\text{and} \quad y < z &&\text{implies} \quad x < z, \\ x < y \quad &\text{and} \quad z > 0 &&\text{implies} \quad xz < yz, \\ x < y \quad &\text{and} \quad x, y > 0 &&\text{implies} \quad \frac{1}{y} < \frac{1}{x}. \end{aligned}$$

We define $x \leqq y$ to mean $x < y$ or $x = y$. Then $x \leqq y$ and $y \leqq x$ imply $x = y$.

If K is ordered and $x \in K$, $x \neq 0$, then x^2 is positive because $x^2 = (-x)^2$ and either $x \in P$ or $-x \in P$. Thus a sum of squares is positive, or 0.

Let E be a field. Then a product of sums of squares in E is a sum of squares. If $a, b \in E$ are sums of squares and $b \neq 0$ then a/b is a sum of squares.

The first assertion is obvious, and the second also, from the expression $a/b = ab(b^{-1})^2$.

If E has characteristic $\neq 2$, and -1 is a sum of squares in E, then every element $a \in E$ is a sum of squares, because $4a = (1 + a)^2 - (1 - a)^2$.

If K is a field with an ordering P, and F is a subfield, then obviously, $P \cap F$ defines an ordering of F, which is called the **induced** ordering.

We observe that our two axioms **ORD 1** and **ORD 2** apply to a ring. If A is an ordered ring, with $1 \neq 0$, then clearly A cannot have divisors of 0, and one can extend the ordering of A to the quotient field in the obvious way: A faction is called positive if it can be written in the form a/b with $a, b \in A$ and $a, b > 0$. One verifies trivially that this defines an ordering on the quotient field.

Example. We define an ordering on the polynomial ring $\mathbf{R}[t]$ over the real numbers. A polynomial

$$f(t) = a_n t^n + \cdots + a_0$$

with $a_n \neq 0$ is defined to be positive if $a_n > 0$. The two axioms are then trivially verified. We note that $t > a$ for all $a \in \mathbf{R}$. Thus t is infinitely large with respect to $\mathbf{R}$. The existence of infinitely large (or infinitely small) elements in an ordered field is the main aspect in which such a field differs from a subfield of the real numbers.

We shall now make some comment on this behavior, i.e. the existence of infinitely large elements.

Let K be an ordered field and let F be a subfield with the induced ordering. As usual, we put $|x| = x$ if $x > 0$ and $|x| = -x$ if $x < 0$. We say that an element α in K is **infinitely large** over F if $|\alpha| \geqq x$ for all $x \in F$. We say that it is **infinitely small** over F if $0 \leqq |\alpha| < |x|$ for all $x \in F, x \neq 0$. We see that α is infinitely large if and only if α^{-1} is infinitely small. We say that K is **archimedean** over F if K has no elements which are infinitely large over F. An intermediate field F_1, $K \supset F_1 \supset F$, is **maximal archimedean over** F in K if it is archimedean over F, and no other intermediate field containing F_1 is archimedean over F. If F_1 is archimedean over F and F_2 is archimedean over F_1 then F_2 is archimedean over F. Hence by Zorn's lemma there always exists a maximal archimedean subfield F_1 of K over F. We say that F is **maximal archimedean in** K if it is maximal archimedean over itself in K.

Let K be an ordered field and F a subfield. Let $\mathfrak{o}$ be the set of elements of K which are not infinitely large over F. Then it is clear that $\mathfrak{o}$ is a ring, and that for any $\alpha \in K$, we have α or $\alpha^{-1} \in \mathfrak{o}$. Hence $\mathfrak{o}$ is what is called a valuation ring, containing F. Let $\mathfrak{m}$ be the ideal of all $\alpha \in K$ which are infinitely small over F. Then $\mathfrak{m}$ is the unique maximal ideal of $\mathfrak{o}$, because any element in $\mathfrak{o}$ which is not in $\mathfrak{m}$ has an inverse in $\mathfrak{o}$. We call $\mathfrak{o}$ the **valuation ring determined by the ordering of K/F.**

Proposition 1.1. *Let K be an ordered field and F a subfield. Let $\mathfrak{o}$ be the valuation ring determined by the ordering of K/F, and let $\mathfrak{m}$ be its maximal ideal. Then $\mathfrak{o}/\mathfrak{m}$ is a real field.*

Proof. Otherwise, we could write

$$-1 = \sum \alpha_i^2 + a$$

with $\alpha_i \in \mathfrak{o}$ and $a \in \mathfrak{m}$. Since $\sum \alpha_i^2$ is positive and a is infinitely small, such a relation is clearly impossible.

§2. REAL FIELDS

A field K is said to be **real** if -1 is not a sum of squares in K. A field K is said to be **real closed** if it is real, and if any algebraic extension of K which is real must be equal to K. In other words, K is maximal with respect to the property of reality in an algebraic closure.

Proposition 2.1. *Let K be a real field.*

(i) *If $a \in K$, then $K(\sqrt{a})$ or $K(\sqrt{-a})$ is real. If a is a sum of squares in K, then $K(\sqrt{a})$ is real. If $K(\sqrt{a})$ is not real, then $-a$ is a sum of squares in K.*

(ii) *If f is an irreducible polynomial of odd degree n in $K[X]$ and if α is a root of f, then $K(\alpha)$ is real.*

Proof. Let $a \in K$. If a is a square in K, then $K(\sqrt{a}) = K$ and hence is real by assumption. Assume that a is not a square in K. If $K(\sqrt{a})$ is not real, then there exist $b_i, c_i \in K$ such that

$$\begin{aligned} -1 &= \sum (b_i + c_i\sqrt{a})^2 \\ &= \sum (b_i^2 + 2c_i b_i\sqrt{a} + c_i^2 a). \end{aligned}$$

Since $\sqrt{a}$ is of degree 2 over K, it follows that

$$-1 = \sum b_i^2 + a \sum c_i^2.$$

If a is a sum of squares in K, this yields a contradiction. In any case, we conclude that

$$-a = \frac{1 + \sum b_i^2}{\sum c_i^2}$$

is a quotient of sums of squares, and by a previous remark, that $-a$ is a sum of squares. Hence $K(\sqrt{a})$ is real, thereby proving our first assertion.

As to the second, suppose $K(\alpha)$ is not real. Then we can write

$$-1 = \sum g_i(\alpha)^2$$

with polynomials g_i in $K[X]$ of degree $\leqq n - 1$. There exists a polynomial h in $K[X]$ such that

$$-1 = \sum g_i(X)^2 + h(X)f(X).$$

The sum of $g_i(X)^2$ has even degree, and this degree must be > 0, otherwise -1 is a sum of squares in K. This degree is $\leqq 2n - 2$. Since f has odd degree n, it follows that h has odd degree $\leqq n - 2$. If β is a root of h then we see that -1 is a sum of squares in $K(\beta)$. Since deg $h <$ deg f, our proof is finished by induction.

Let K be a real field. By a **real closure** we shall mean a real closed field L which is algebraic over K.

Theorem 2.2. *Let K be a real field. Then there exists a real closure of K. If R is real closed, then R has a unique ordering. The positive elements are the squares of R. Every positive element is a square, and every polynomial of odd degree in $R[X]$ has a root in R. We have $R^a = R(\sqrt{-1})$.*

Proof. By Zorn's lemma, our field K is contained in some real closed field algebraic over K. Now let R be a real closed field. Let P be the set of non-zero elements of R which are sums of squares. Then P is closed under addition and multiplication. By Proposition 2.1, every element of P is a square in R, and given $a \in R$, $a \neq 0$, we must have $a \in P$ or $-a \in P$. Thus P defines an ordering. Again by Proposition 2.1, every polynomial of odd degree over R has a root in R. Our assertion follows by Example 5 of Chapter VI, §2.

Corollary 2.3. *Let K be a real field and a an element of K which is not a sum of squares. Then there exists an ordering of K in which a is negative.*

Proof. The field $K(\sqrt{-a})$ is real by Proposition 1.1 and hence has an ordering as a subfield of a real closure. In this ordering, $-a > 0$ and hence a is negative.

Proposition 2.4. *Let R be a field such that $R \neq R^a$ but $R^a = R(\sqrt{-1})$. Then R is real and hence real closed.*

Proof. Let P be the set of elements of R which are squares and $\neq 0$. We contend that P is an ordering of R. Let $a \in R$, $a \neq 0$. Suppose that a is not a square in R. Let α be a root of $X^2 - a = 0$. Then $R(\alpha) = R(\sqrt{-1})$, and hence there exist $c, d \in R$ such that $\alpha = c + d\sqrt{-1}$. Then

$$\alpha^2 = c^2 + 2cd\sqrt{-1} - d^2.$$

Since $1, \sqrt{-1}$ are linearly independent over R, it follows that $c = 0$ (because $a \notin R^2$), and hence $-a$ is a square.

We shall now prove that a sum of squares is a square. For simplicity, write $i = \sqrt{-1}$. Since $R(i)$ is algebraically closed, given $a, b \in R$ we can find $c, d \in R$ such that $(c + di)^2 = a + bi$. Then $a = c^2 - d^2$ and $b = 2cd$. Hence

$$a^2 + b^2 = (c^2 + d^2)^2,$$

as was to be shown.

If $a \in R$, $a \neq 0$, then not both a and $-a$ can be squares in R. Hence P is an ordering and our proposition is proved.

Theorem 2.5. *Let R be a real closed field, and $f(X)$ a polynomial in $R[X]$. Let $a, b \in R$ and assume that $f(a) < 0$ and $f(b) > 0$. Then there exists c between a and b such that $f(c) = 0$.*

Proof. Since $R(\sqrt{-1})$ is algebraically closed, it follows that f splits into a product of irreducible factors of degree 1 or 2. If $X^2 + \alpha X + \beta$ is irreducible ($\alpha, \beta \in R$) then it is a sum of squares, namely

$$\left(X + \frac{\alpha}{2}\right)^2 + \left(\beta - \frac{\alpha^2}{4}\right),$$

and we must have $4\beta > \alpha^2$ since our factor is assumed irreducible. Hence the change of sign of f must be due to the change of sign of a linear factor, which is trivially verified to be a root lying between a and b.

Lemma 2.6. *Let K be a subfield of an ordered field E. Let $\alpha \in E$ be algebraic over K, and a root of the polynomial*

$$f(X) = X^n + a_{n-1}X^{n-1} + \cdots + a_0$$

with coefficients in K. Then $|\alpha| \leqq 1 + |a_{n-1}| + \cdots + |a_0|$.

Proof. If $|\alpha| \leqq 1$, the assertion is obvious. If $|\alpha| > 1$, we express $|\alpha|^n$ in terms of the terms of lower degree, divide by $|\alpha|^{n-1}$, and get a proof for our lemma.

Note that the lemma implies that an element which is algebraic over an ordered field cannot be infinitely large with respect to that field.

Let $f(X)$ be a polynomial with coefficients in a real closed field R, and assume that f has no multiple roots. Let $u < v$ be elements of R. By a **Sturm sequence** for f over the interval $[u, v]$ we shall mean a sequence of polynomials

$$S = \{f = f_0, f' = f_1, \ldots, f_m\}$$

having the following properties:

ST 1. The last polynomial f_m is a non-zero constant.

ST 2. There is no point $x \in [u, v]$ such that $f_j(x) = f_{j+1}(x) = 0$ for any value $0 \leqq j \leqq m - 1$.

ST 3. If $x \in [u, v]$ and $f_j(x) = 0$ for some $j = 1, \ldots, m - 1$, then $f_{j-1}(x)$ and $f_{j+1}(x)$ have opposite signs.

ST 4. We have $f_j(u) \neq 0$ and $f_j(v) \neq 0$ for all $j = 0, \ldots, m$.

For any $x \in [u, v]$ which is not a root of any polynomial f_i we denote by $W_S(x)$ the number of sign changes in the sequence

$$\{f(x), f_1(x), \ldots, f_m(x)\},$$

and call $W_S(x)$ the **variation of signs in the sequence**.

Theorem 2.7. (Sturm's Theorem). *The number of roots of f between u and v is equal to $W_S(u) - W_S(v)$ for any Sturm sequence S.*

Proof. We observe that if $\alpha_1 < \alpha_2 < \cdots < \alpha_r$ is the ordered sequence of roots of the polynomials f_j in $[u, v]$ ($j = 0, \ldots, m - 1$), then $W_S(x)$ is constant on the open intervals between these roots, by Theorem 2.5. Hence it will suffice to prove that if there is precisely one element α such that $u < \alpha < v$ and α is a root of some f_j, then $W_S(u) - W_S(v) = 1$ if α is a root of f, and 0 otherwise. Suppose that α is a root of some f_j, for $1 \leqq j \leqq m - 1$. Then $f_{j-1}(\alpha)$, $f_{j+1}(\alpha)$ have opposite signs by **ST 3**, and these signs do not change when we replace α by u or v. Hence the variation of signs in

$$\{f_{j-1}(u), f_j(u), f_{j+1}(u)\} \quad \text{and} \quad \{f_{j-1}(v), f_j(v), f_{j+1}(v)\}$$

is the same, namely equal to 2. If α is not a root of f, we conclude that

$$W_S(u) = W_S(v).$$

If α is a root of f, then $f(u)$ and $f(v)$ have opposite signs, but $f'(u)$ and $f'(v)$ have the same sign, namely, the sign of $f'(\alpha)$. Hence in this case,

$$W_S(u) = W_S(v) + 1.$$

This proves our theorem.

It is easy to construct a Sturm sequence for a polynomial without multiple roots. We use the Euclidean algorithm, writing

$$\begin{aligned} f &= g_1 f' - f_2, \\ f_2 &= g_2 f_1 - f_3, \\ &\vdots \\ f_{m-2} &= g_{m-1} f_{m-1} - f_m, \end{aligned}$$

using $f' = f_1$. Since f, f' have no common factor, the last term of this sequence is non-zero constant. The other properties of a Sturm sequence are trivially verified, because if two successive polynomials of the sequence have a common zero, then they must all be 0, contradicting the fact that the last one is not.

Corollary 2.8. *Let K be an ordered field, f an irreducible polynomial of degree $\geqq 1$ over K. The number of roots of f in two real closures of K inducing the given ordering on K is the same.*

Proof. We can take v sufficiently large positive and u sufficiently large negative in K so that all roots of f and all roots of the polynomials in the Sturm sequence lie between u and v, using Lemma 2.6. Then $W_S(u) - W_S(v)$ is the total number of roots of f in any real closure of K inducing the given ordering.

Theorem 2.9. *Let K be an ordered field, and let R, R' be real closures of K, whose orderings induce the given ordering on K. Then there exists a unique isomorphism $\sigma: R \to R'$ over K, and this isomorphism is order-preserving.*

Proof. We first show that given a finite subextension E of R over K, there exists an embedding of E into R' over K. Let $E = K(\alpha)$, and let

$$f(X) = \text{Irr}(\alpha, K, X).$$

Then $f(\alpha) = 0$ and the corollary of Sturm's Theorem (Corollary 2.8) shows that f has a root β in R'. Thus there exists an isomorphism of $K(\alpha)$ on $K(\beta)$ over K, mapping α on β.

Let $\alpha_1, \dots, \alpha_n$ be the distinct roots of f in R, and let $\beta_1, \dots, \beta_m$ be the distinct roots of f in R'. Say

$$\alpha_1 < \cdots < \alpha_n \quad \text{in the ordering of } R,$$
$$\beta_1 < \cdots < \beta_m \quad \text{in the ordering of } R'.$$

We contend that $m = n$ and that we can select an embedding σ of $K(\alpha_1, \dots, \alpha_n)$ into R' such that $\sigma\alpha_i = \beta_i$ for $i = 1, \dots, n$. Indeed, let γ_i be an element of R such that

$$\gamma_i^2 = \alpha_{i+1} - \alpha_i \quad \text{for} \quad i = 1, \dots, n-1$$

and let $E_1 = K(\alpha_1, \dots, \alpha_n, \gamma_1, \dots, \gamma_{n-1})$. By what we have seen, there exists an embedding σ of E_1 into R', and then $\sigma\alpha_{i+1} - \sigma\alpha_i$ is a square in R'. Hence

$$\sigma\alpha_1 < \cdots < \sigma\alpha_n.$$

This proves that $m \geqq n$. By symmetry, it follows that $m = n$. Furthermore, the condition that $\sigma\alpha_i = \beta_i$ for $i = 1, \dots, n$ determines the effect of σ on

$K(\alpha_1, \ldots, \alpha_n)$. We contend that σ is order-preserving. Let $y \in K(\alpha_1, \ldots, \alpha_n)$ and $0 < y$. Let $\gamma \in R$ be such that $\gamma^2 = y$. There exists an embedding of

$$K(\alpha_1, \ldots, \alpha_n, \gamma_1, \ldots, \gamma_{n-1}, \gamma)$$

into R' over K which must induce σ on $K(\alpha_1, \ldots, \alpha_n)$ and is such that σy is a square, hence > 0, as contended.

Using Zorn's lemma, it is now clear that we get an isomorphism of R onto R' over K. This isomorphism is order-preserving because it maps squares on squares, thereby proving our theorem.

Proposition 2.10. *Let K be an ordered field, K' an extension such that there is no relation*

$$-1 = \sum_{i=1}^{n} a_i \alpha_i^2$$

with $a_i \in K$, $a_i > 0$, and $\alpha_i \in K'$. Let L be the field obtained from K' by adjoining the square roots of all positive elements of K. Then L is real.

Proof. If not, there exists a relation of type

$$-1 = \sum_{i=1}^{n} a_i \alpha_i^2$$

with $a_i \in K$, $a_i > 0$, and $\alpha_i \in L$. (We can take $a_i = 1$.) Let r be the smallest integer such that we can write such a relation with α_i in a subfield of L, of type

$$K'(\sqrt{b_1}, \ldots, \sqrt{b_r})$$

with $b_j \in K$, $b_j > 0$. Write

$$\alpha_i = x_i + y_i\sqrt{b_r}$$

with $x_i, y_i \in K'(\sqrt{b_1}, \ldots, \sqrt{b_{r-1}})$. Then

$$\begin{aligned} -1 &= \sum a_i(x_i + y_i\sqrt{b_r})^2 \\ &= \sum a_i(x_i^2 + 2x_i y_i\sqrt{b_r} + y_i^2 b_r). \end{aligned}$$

By hypothesis, $\sqrt{b_r}$ is not in $K'(b_1, \ldots, \sqrt{b_{r-1}})$. Hence

$$-1 = \sum a_i x_i^2 + \sum a_i b_r y_i^2,$$

contradicting the minimality of r.

Theorem 2.11. *Let K be an ordered field. There exists a real closure R of K inducing the given ordering on K.*

Proof. Take $K' = K$ in Proposition 2.10. Then L is real, and is contained in a real closure. Our assertion is clear.

Corollary 2.12. *Let K be an ordered field, and K' an extension field. In order that there exist an ordering on K' inducing the given ordering of K, it is necessary and sufficient that there is no relation of type*

$$-1 = \sum_{i=1}^{n} a_i \alpha_i^2$$

with $a_i \in K$, $a_i > 0$, and $\alpha_i \in K'$.

Proof. If there is no such relation, then Proposition 2.10 states that L is contained in a real closure, whose ordering induces an ordering on K', and the given ordering on K, as desired. The converse is clear.

Example. Let $\mathbf{Q}^a$ be the field of algebraic numbers. One sees at once that $\mathbf{Q}$ admits only one ordering, the ordinary one. Hence any two real closures of $\mathbf{Q}$ in $\mathbf{Q}^a$ are isomorphic, by means of a unique isomorphism. The real closures of $\mathbf{Q}$ in $\mathbf{Q}^a$ are precisely those subfields of $\mathbf{Q}^a$ which are of finite degree under $\mathbf{Q}^a$. Let K be a finite real extension of $\mathbf{Q}$, contained in $\mathbf{Q}^a$. An element α of K is a sum of squares in K if and only if every conjugate of α in the real numbers is positive, or equivalently, if and only if every conjugate of α in one of the real closures of $\mathbf{Q}$ in $\mathbf{Q}^a$ is positive.

Note. The theory developed in this and the preceding section is due to Artin-Schreier. See the bibliography at the end of the chapter.

§3. REAL ZEROS AND HOMOMORPHISMS

Just as we developed a theory of extension of homomorphisms into an algebraically closed field, and Hilbert's Nullstellensatz for zeros in an algebraically closed field, we wish to develop the theory for values in a real closed field. One of the main theorems is the following:

Theorem 3.1. *Let k be a field, $K = k(x_1, \ldots, x_n)$ a finitely generated extension. Assume that K is ordered. Let R_k be a real closure of k inducing the same ordering on k as K. Then there exists a homomorphism*

$$\varphi : k[x_1, \ldots, x_n] \to R_k$$

over k.

As applications of Theorem 3.1, one gets:

Corollary 3.2. *Notation being as in the theorem, let* $y_1, \dots, y_m \in k[x]$ *and assume*

$$y_1 < y_2 < \cdots < y_m$$

is the given ordering of K. *Then one can choose* φ *such that*

$$\varphi y_1 < \cdots < \varphi y_m.$$

Proof. Let $\gamma_i \in K^a$ be such that $\gamma_i^2 = y_{i+1} - y_i$. Then $K(\gamma_1, \dots, \gamma_{n-1})$ has an ordering inducing the given ordering on K. We apply the theorem to the ring

$$k[x_1, \dots, x_n, \gamma_1^{-1}, \dots, \gamma_{m-1}^{-1}, \gamma_1, \dots, \gamma_{m-1}].$$

Corollary 3.3. (Artin). *Let k be a real field admitting only one ordering. Let $f(X_1, \dots, X_n) \in k(X)$ be a rational function having the property that for all $(a) = (a_1, \dots, a_n) \in R_k^{(n)}$ such that $f(a)$ is defined, we have $f(a) \geqq 0$. Then $f(X)$ is a sum of squares in $k(X)$.*

Proof. Assume that our conclusion is false. By Corollary 2.3, there exists an ordering of $k(X)$ in which f is negative. Apply Corollary 3.2 to the ring

$$k[X_1, \dots, X_n, h(X)^{-1}]$$

where $h(X)$ is a polynomial denominator for $f(X)$. We can find a homomorphism φ of this ring into R_k (inducing the identity on k) such that $\varphi(f) < 0$. But

$$\varphi(f) = f(\varphi X_1, \dots, \varphi X_n).$$

contradiction. We let $a_i = \varphi(X_i)$ to conclude the proof.

Corollary 3.3 was a Hilbert problem. The proof which we shall describe for Theorem 3.1 differs from Artin's proof of the corollary in several technical aspects.

We shall first see how one can reduce Theorem 3.1 to the case when K has transcendence degree 1 over k, and k is real closed.

Lemma 3.4. *Let R be a real closed field and let R_0 be a subfield which is algebraically closed in R (i.e. such that every element of R not in R_0 is transcendental over R_0). Then R_0 is real closed.*

Proof. Let $f(X)$ be an irreducible polynomial over R_0. It splits in R into linear and quadratic factors. Its coefficients in R are algebraic over R_0, and hence must lie in R_0. Hence $f(X)$ is linear itself, or quadratic irreducible already over R_0. By the intermediate value theorem, we may assume that f is positive

definite, i.e. $f(a) > 0$ for all $a \in R_0$. Without loss of generality, we may assume that $f(X) = X^2 + b^2$ for some $b \in R_0$. Any root of this polynomial will bring $\sqrt{-1}$ with it and therefore the only algebraic extension of R_0 is $R_0(\sqrt{-1})$. This proves that R_0 is real closed.

Let R_K be a real closure of K inducing the given ordering on K. Let R_0 be the algebraic closure of k in R_K. By the lemma, R_0 is real closed.

We consider the field $R_0(x_1, \ldots, x_n)$. If we can prove our theorem for the ring $R_0[x_1, \ldots, x_n]$, and find a homomorphism

$$\psi : R_0[x_1, \ldots, x_n] \to R_0,$$

then we let $\sigma : R_0 \to R_K$ be an isomorphism over k (it exists by Theorem 2.9), and we let $\varphi = \sigma \circ \psi$ to solve our problem over k. This reduces our theorem to the case when k is real closed.

Next, let F be an intermediate field, $K \supset F \supset k$, such that K is of transcendence degree 1 over F. Again let R_K be a real closure of K preserving the ordering, and let R_F be the real closure of F contained in R_K. If we know our theorem for extensions of dimension 1, then we can find a homomorphism

$$\psi : R_F[x_1, \ldots, x_n] \to R_F.$$

We note that the field $k(\psi x_1, \ldots, \psi x_n)$ has transcendence degree $\leqq n - 1$, and is real, because it is contained in R_F. Thus we are reduced inductively to the case when K has dimension 1, and as we saw above, when k is real closed.

One can interpret our statement geometrically as follows. We can write $K = R(x, y)$ with x transcendental over R, and (x, y) satisfying some irreducible polynomial $f(X, Y) = 0$ in $R[X, Y]$. What we essentially want to prove is that there are infinitely many points on the curve $f(X, Y) = 0$, with coordinates lying in R, i.e. infinitely many real points.

The main idea is that we find some point $(a, b) \in R^{(2)}$ such that $f(a, b) = 0$ but $D_2 f(a, b) \neq 0$. We can then use the intermediate value theorem. We see that $f(a, b + h)$ changes sign as h changes from a small positive to a small negative element of R. If we take $a' \in R$ close to a, then $f(a', b + h)$ also changes sign for small h, and hence $f(a', Y)$ has a zero in R for all a' sufficiently close to a. In this way we get infinitely many zeros.

To find our point, we consider the polynomial $f(x, Y)$ as a polynomial in one variable Y with coefficients in $R(x)$. Without loss of generality we may assume that this polynomial has leading coefficient 1. We construct a Sturm sequence for this polynomial, say

$$\{f(x, Y), f_1(x, Y), \ldots, f_m(x, Y)\}.$$

Let $d = \deg f$. If we denote by $A(x) = (a_{d-1}(x), \ldots, a_0(x))$ the coefficients of $f(x, Y)$, then from the Euclidean alogrithm, we see that the coefficients of the

polynomials in the Sturm sequence can be expressed as rational functions

$$\{G_\nu(A(x))\}$$

in terms of $a_{d-1}(x), \ldots, a_0(x)$.

Let

$$v(x) = 1 \pm a_{d-1}(x) \pm \cdots \pm a_0(x) + s,$$

where s is a positive integer, and the signs are selected so that each term in this sum gives a positive contribution. We let $u(x) = -v(x)$, and select s so that neither u nor v is a root of any polynomial in the Sturm sequence for f. Now we need a lemma.

Lemma 3.5. *Let R be a real closed field, and $\{h_i(x)\}$ a finite set of rational functions in one variable with coefficients in R. Suppose the rational field $R(x)$ ordered in some way, so that each $h_i(x)$ has a sign attached to it. Then there exist infinitely many special values c of x in R such that $h_i(c)$ is defined and has the same sign as $h_i(x)$, for all i.*

Proof. Considering the numerators and denominators of the rational functions, we may assume without loss of generality that the h_i are polynomials. We then write

$$h_i(x) = \alpha \prod (x - \lambda) \prod p(x),$$

where the first product is extended over all roots λ of h_i in R, and the second product is over positive definite quadratic factors over R. For any $\xi \in R$, $p(\xi)$ is positive. It suffices therefore to show that the signs of $(x - \lambda)$ can be preserved for all λ by substituting infinitely many values α for x. We order all values of λ and of x and obtain

$$\cdots < \lambda_1 < x < \lambda_2 < \cdots$$

where possibly λ_1 or λ_2 is omitted if x is larger or smaller than any λ. Any value α of x in R selected between λ_1 and λ_2 will then satisfy the requirements of our lemma.

To apply the lemma to the existence of our point, we let the rational functions $\{h_1(x)\}$ consist of all coefficients $a_{d-1}(x), \ldots, a_0(x)$, all rational functions $G_\nu(A(x))$, and all values $f_j(x, u(x))$, $f_j(x, v(x))$ whose variation in signs satisfied Sturm's theorem. We then find infinitely many special values α of x in R which preserve the signs of these rational functions. Then the polynomials $f(\alpha, Y)$ have roots in R, and for all but a finite number of α, these roots have multiplicity 1.

It is then a matter of simple technique to see that for all but a finite number of points on the curve, the elements $x_1, \ldots, x_n$ lie in the local ring of the homomorphism $R[x, y] \to R$ mapping (x, y) on (a, b) such that $f(a, b) = 0$ but

$D_2 f(a, b) \neq 0$. (Cf. for instance the example at the end of §4, Chapter XII, and Exercise 18 of that chapter.) One could also give direct proofs here. In this way, we obtain homomorphisms

$$R[x_1, \dots, x_n] \to R,$$

thereby proving Theorem 3.1.

Theorem 3.6. *Let k be a real field, $K = k(x_1, \dots, x_n, y) = k(x, y)$ a finitely generated extension such that $x_1, \dots, x_n$ are algebraically independent over k, and y is algebraic over $k(x)$. Let $f(X, Y)$ be the irreducible polynomial in $k[X, Y]$ such that $f(x, y) = 0$. Let R be a real closed field containing k, and assume that there exists $(a, b) \in R^{(n+1)}$ such that $f(a, b) = 0$ but*

$$D_{n+1} f(a, b) \neq 0.$$

Then K is real.

Proof. Let $t_1, \dots, t_n$ be algebraically independent over R. Inductively, we can put an ordering on $R(t_1, \dots, t_n)$ such that each t_i is infinitely small with respect to R, (cf. the example in §1). Let R' be a real closure of $R(t_1, \dots, t_n)$ preserving the ordering. Let $u_i = a_i + t_i$ for each $i = 1, \dots, n$. Then $f(u, b + h)$ changes sign for small h positive and negative in R, and hence $f(u, Y)$ has a root in R', say v. Since f is irreducible, the isomorphism of $k(x)$ on $k(u)$ sending x_i on u_i extends to an embedding of $k(x, y)$ into R', and hence K is real, as was to be shown.

In the language of algebraic geometry, Theorems 3.1 and 3.6 state that the function field of a variety over a real field k is real if and only if the variety has a simple point in some real closure of k.

EXERCISES

1. Let α be algebraic over $\mathbf{Q}$ and assume that $\mathbf{Q}(\alpha)$ is a real field. Prove that α is a sum of squares in $\mathbf{Q}(\alpha)$ if and only if for every embedding σ of $\mathbf{Q}(\alpha)$ in $\mathbf{R}$ we have $\sigma\alpha > 0$.
2. Let F be a finite extension of $\mathbf{Q}$. Let $\varphi : F \to \mathbf{Q}$ be a $\mathbf{Q}$-linear functional such that $\varphi(x^2) > 0$ for all $x \in F, x \neq 0$. Let $\alpha \in F, \alpha \neq 0$. If $\varphi(\alpha x^2) \geqq 0$ for all $x \in F$, show that α is a sum of squares in F, and that F is totally real, i.e. every embedding of F in the complex numbers is contained in the real numbers. [*Hint*: Use the fact that the trace gives an identification of F with its dual space over $\mathbf{Q}$, and use the approximation theorem of Chapter XII, §1.]

3. Let $\alpha \leqq t \leqq \beta$ be a real interval, and let $f(t)$ be a real polynomial which is positive on this interval. Show that $f(t)$ can be written in the form

$$c\left(\sum Q_\nu^2 + \sum (t-\alpha)Q_\mu^2 + \sum (\beta - t)Q_\lambda^2\right)$$

where Q^2 denotes a square, and $c \geqq 0$. *Hint*: Split the polynomial, and use the identity:

$$(t-\alpha)(\beta-t) = \frac{(t-\alpha)^2(\beta-t) + (t-\alpha)(\beta-t)^2}{\beta-\alpha}.$$

Remark. The above seemingly innocuous result is a key step in developing the spectral theorem for bounded hermitian operators on Hilbert space. See the appendix of [La 72] and also [La 85].

4. Show that the field of real numbers has only the identity automorphism. [*Hint*: Show that an automorphism preserves the ordering.]

Real places

For the next exercises, cf. Krull [Kr 32] and Lang [La 53]. These exercises form a connected sequence, and solutions will be found in [La 53].

5. Let K be a field and suppose that there exists a real place of K; that is, a place φ with values in a real field L. Show that K is real.

6. Let K be an ordered real field and let F be a subfield which is maximal archimedean in K. Show that the canonical place of K with respect to F is algebraic over F (i.e. if $\mathfrak{o}$ is the valuation ring of elements of K which are not infinitely large over F, and $\mathfrak{m}$ is its maximal ideal, then $\mathfrak{o}/\mathfrak{m}$ is algebraic over F).

7. Let K be an ordered field and let F be a subfield which is maximal archimedean in K. Let K' be the real closure of K (preserving the ordering), and let F' be the real closure of F contained in K'. Let φ be the canonical place of K' with respect to F'. Show that $\varphi(K')$ is F'-valued, and that the restriction of φ to K is equivalent to the canonical place of K over F.

8. Define a real field K to be **quadratically closed** if for all $\alpha \in K$ either $\sqrt{\alpha}$ or $\sqrt{-\alpha}$ lies in K. The ordering of a quadratically closed real field K is then uniquely determined, and so is the real closure of such a field, up to an isomorphism over K. Suppose that K is quadratically closed. Let F be a subfield of K and suppose that F is maximal archimedean in K. Let φ be a place of K over F, with values in a field which is algebraic over F. Show that φ is equivalent to the canonical place of K over F.

9. Let K be a quadratically closed real field. Let φ be a real place of K, taking its values in a real closed field R. Let F be a maximal subfield of K such that φ is an isomorphism on F, and identify F with $\varphi(F)$. Show that such F exists and is maximal archimedean in K. Show that the image of φ is algebraic over F, and that φ is induced by the canonical place of K over F.

10. Let K be a real field and let φ be a real place of K, taking its values in a real closed field R. Show that there is an extension of φ to an R-valued place of a real closure of K. [*Hint*: first extend φ to a quadratic closure of K. Then use Exercise 5.]

11. Let $K \subset K_1 \subset K_2$ be real closed fields. Suppose that K is maximal archimedean in K_1 and K_1 is maximal archimedean in K_2. Show that K is maximal archimedean in K_2.

12. Let K be a real closed field. Show that there exists a real closed field R containing K and having arbitrarily large transcendence degree over K, and such that K is maximal archimedean in R.

13. Let R be a real closed field. Let $f_1, \ldots, f_r$ be homogeneous polynomials of odd degrees in n variables over R. If $n > r$, show that these polynomials have a non-trivial common zero in R. (*Comments*: If the forms are generic (in the sense of Chapter IX), and $n = r + 1$, it is a theorem of Bezout that in the algebraic closure R^a the forms have exactly $d_1 \cdots d_m$ common zeros, where d_i is the degree of f_i. You may assume this to prove the result as stated. If you want to see this worked out, see [La 53], Theorem 15. Compare with Exercise 3 of Chapter IX.)

Bibliography

[Ar 24] E. Artin, Kennzeichnung des Körpers der reellen algebraischen Zahlen, *Abh. Math. Sem. Hansischen Univ.* **3** (1924), pp. 319–323

[Ar 27] E. Artin, Über die Zerlegung definiter Funktionen in Quadrate, *Abh. Math. Sem. Hansischen Univ.* **5** (1927), pp. 100–115

[ArS 27] E. Artin and E. Schreier, Algebraische Konstruktion reeller Körper, *Abh. Math. Sem. Hansischen Univ.* **5** (1927), pp. 85–99

[Kr 32] W. Krull, Allgemeine Bewertungstheorie, *J. reine angew. Math.* (1932), pp. 169–196

[La 53] S. Lang, The theory of real places, *Ann. Math.* **57** No. 2 (1953), pp. 378–391

[La 72] S. Lang, *Differential manifolds*, Addison-Wesley, 1972; reprinted by Springer Verlag, 1985; superceded by [La 99a].

[La 85] S. Lang, *Real and functional analysis*. Third edition, Springer Verlag, 1993

[La 99a] S. Lang, *Fundamentals of Differential Geometry*, Springer Verlag, 1999

CHAPTER XII

Absolute Values

§1. DEFINITIONS, DEPENDENCE, AND INDEPENDENCE

Let K be a field. An **absolute value** v on K is a real-valued function $x \mapsto |x|_v$ on K satisfying the following three properties:

AV 1. We have $|x|_v \geqq 0$ for all $x \in K$, and $|x|_v = 0$ if and only if $x = 0$.

AV 2. For all $x, y \in K$, we have $|xy|_v = |x|_v |y|_v$.

AV 3. For all $x, y \in K$, we have $|x + y|_v \leqq |x|_v + |y|_v$.

If instead of **AV 3** the absolute value satisfies the stronger condition

AV 4. $|x + y|_v \leqq \max(|x|_v, |y|_v)$

then we shall say that it is a **valuation**, or that it is non-archimedean.

The absolute value which is such that $|x|_v = 1$ for all $x \neq 0$ is called **trivial**.

We shall write $|x|$ instead of $|x|_v$ if we deal with just one fixed absolute value. We also refer to v as the absolute value.

An absolute value of K defines a metric. The distance between two elements x, y of K in this metric is $|x - y|$. Thus an absolute value defines a topology on K. Two absolute values are called **dependent** if they define the same topology. If they do not, they are called independent.

We observe that $|1| = |1^2| = |(-1)^2| = |1|^2$ whence

$$|1| = |-1| = 1.$$

Also, $|-x| = |x|$ for all $x \in K$, and $|x^{-1}| = |x|^{-1}$ for $x \neq 0$.

Proposition 1.1. *Let $|\ |_1$ and $|\ |_2$ be non-trivial absolute values on a field K. They are dependent if and only if the relation*

$$|x|_1 < 1$$

implies $|x|_2 < 1$. If they are dependent, then there exists a number $\lambda > 0$ such that $|x|_1 = |x|_2^\lambda$ for all $x \in K$.

Proof. If the two absolute values are dependent, then our condition is satisfied, because the set of $x \in K$ such that $|x|_1 < 1$ is the same as the set such that $\lim x^n = 0$ for $n \to \infty$. Conversely, assume the condition satisfied. Then $|x|_1 > 1$ implies $|x|_2 > 1$ since $|x^{-1}|_1 < 1$. By hypothesis, there exists an element $x_0 \in K$ such that $|x_0|_1 > 1$. Let $a = |x_0|_1$ and $b = |x_0|_2$. Let

$$\lambda = \frac{\log b}{\log a}.$$

Let $x \in K, x \neq 0$. Then $|x|_1 = |x_0|_1^\alpha$ for some number α. If m, n are integers such that $m/n > \alpha$ and $n > 0$, we have

$$|x|_1 > |x_0|_1^{m/n}$$

whence

$$|x^n/x_0^m|_1 < 1,$$

and thus

$$|x^n/x_0^m|_2 < 1.$$

This implies that $|x|_2 < |x_0|_2^{m/n}$. Hence

$$|x|_2 \leqq |x_0|_2^\alpha.$$

Similarly, one proves the reverse inequality, and thus one gets

$$|x|_2 = |x_0|_2^\alpha$$

for all $x \in K$, $x \neq 0$. The assertion of the proposition is now obvious, i.e. $|x|_2 = |x|_1^\lambda$.

We shall give some examples of absolute values.

Consider first the rational numbers. We have the ordinary absolute value such that $|m| = m$ for any positive integer m.

For each prime number p, we have the p-adic absolute value v_p, defined by the formula

$$|p^r m/n|_p = 1/p^r$$

where r is an integer, and m, n are integers $\neq 0$, not divisible by p. One sees at once that the p-adic absolute value is non-archimedean.

One can give a similar definition of a valuation for any field K which is the quotient field of a principal ring. For instance, let $K = k(t)$ where k is a field and t is a variable over k. We have a valuation v_p for each irreducible polynomial $p(t)$ in $k[t]$, defined as for the rational numbers, but there is no way of normalizing it in a natural way. Thus we select a number c with $0 < c < 1$ and for any rational function $p^r f/g$ where f, g are polynomials not divisible by p, we define

$$|p^r f/g|_p = c^r.$$

The various choices of the constant c give rise to dependent valuations.

Any subfield of the complex numbers (or real numbers) has an absolute value, induced by the ordinary absolute value on the complex numbers. We shall see later how to obtain absolute values on certain fields by embedding them into others which are already endowed with natural absolute values.

Suppose that we have an absolute value on a field which is bounded on the prime ring (i.e. the integers $\mathbf{Z}$ *if the characteristic is* 0, *or the integers* mod p *if the characteristic is* p). *Then the absolute value is necessarily non-archimedean.*

Proof. For any elements x, y and any positive integer n, we have

$$|(x + y)^n| \leqq \sum \left| \binom{n}{\nu} x^\nu y^{n-\nu} \right| \leqq nC \max(|x|, |y|)^n.$$

Taking n-th roots and letting n go to infinity proves our assertion. We note that this is always the case in characteristic > 0 because the prime ring is finite!

If the absolute value is archimedean, then we refer the reader to any other book in which there is a discussion of absolute values for a proof of the fact that it is dependent on the ordinary absolute value. This fact is essentially useless (and is never used in the sequel), because we always start with a concretely given set of absolute values on fields which interest us.

In Proposition 1.1 we derived a strong condition on dependent absolute values. We shall now derive a condition on independent ones.

Theorem 1.2. (Approximation Theorem). (Artin-Whaples). *Let K be a field and $|\ |_1, \ldots, |\ |_s$ non-trivial pairwise independent absolute values on K. Let $x_1, \ldots, x_s$ be elements of K, and $\epsilon > 0$. Then there exists $x \in K$ such that*

$$|x - x_i|_i < \epsilon$$

for all i.

Proof. Consider first two of our absolute values, say v_1 and v_2. By hypothesis we can find $\alpha \in K$ such that $|\alpha|_1 < 1$ and $|\alpha|_s \geqq 1$. Similarly, we can find $\beta \in K$ such that $|\beta|_1 \geqq 1$ and $|\beta|_s < 1$. Put $y = \beta/\alpha$. Then $|y|_1 > 1$ and $|y|_s < 1$.

We shall now prove that there exists $z \in K$ such that $|z|_1 > 1$ and $|z|_j < 1$ for $j = 2, \dots, s$. We prove this by induction, the case $s = 2$ having just been proved. Suppose we have found $z \in K$ satisfying

$$|z|_1 > 1 \quad \text{and} \quad |z|_j < 1 \qquad \text{for} \quad j = 2, \dots, s-1.$$

If $|z|_s \leqq 1$ then the element $z^n y$ for large n will satisfy our requirements.

If $|z|_s > 1$, then the sequence

$$t_n = \frac{z^n}{1 + z^n}$$

tends to 1 at v_1 and v_s, and tends to 0 at $v_j (j = 2, \dots, s-1)$. For large n, it is then clear that $t_n y$ satisfies our requirements.

Using the element z that we have just constructed, we see that the sequence $z^n/(1 + z^n)$ tends to 1 at v_1 and to 0 at v_j for $j = 2, \dots, s$. For each $i = 1, \dots, s$ we can therefore construct an element z_i which is very close to 1 at v_i and very close to 0 at v_j $(j \neq i)$. The element

$$x = z_1 x_1 + \cdots + z_s x_s$$

then satisfies the requirement of the theorem.

§2. COMPLETIONS

Let K be a field with a non-trivial absolute value v, which will remain fixed throughout this section. One can then define in the usual manner the notion of a Cauchy sequence. It is a sequence $\{x_n\}$ of elements in K such that, given $\epsilon > 0$, there exists an integer N such that for all $n, m > N$ we have

$$|x_n - x_m| < \epsilon.$$

We say that K is **complete** if every Cauchy sequence converges.

Proposition 2.1. *There exists a pair (K_v, i) consisting of a field K_v, complete under an absolute value, and an embedding $i: K \to K_v$ such that the absolute value on K is induced by that of K_v (i.e. $|x|_v = |ix|$ for $x \in K$), and such that iK is dense in K_v. If (K'_v, i') is another such pair, then there exists a unique*

isomorphism $\varphi : K_v \to K'_v$ *preserving the absolute values, and making the following diagram commutative:*

$$\begin{array}{ccc} K_v & \xrightarrow{\varphi} & K'_v \\ {}_{i}\nwarrow & & \nearrow_{i'} \\ & K & \end{array}$$

Proof. The uniqueness is obvious. One proves the existence in the well-known manner, which we shall now recall briefly, leaving the details to the reader.

The Cauchy sequences form a ring, addition and multiplication being taken componentwise.

One defines a null sequence to be a sequence $\{x_n\}$ such that $\lim_{n\to\infty} x_n = 0$. The null sequences form an ideal in the ring of Cauchy sequences, and in fact form a maximal ideal. (If a Cauchy sequence is not a null sequence, then it stays away from 0 for all n sufficiently large, and one can then take the inverse of almost all its terms. Up to a finite number of terms, one then gets again a Cauchy sequence.)

The residue class field of Cauchy sequences modulo null sequences is the field K_v. We embed K in K_v "on the diagonal", i.e. send $x \in K$ on the sequence $(x, x, x, \ldots)$.

We extend the absolute value of K to K_v by continuity. If $\{x_n\}$ is a Cauchy sequence, representing an element ξ in K_v, we define $|\xi| = \lim |x_n|$. It is easily proved that this yields an absolute value (independent of the choice of representative sequence $\{x_n\}$ for ξ), and this absolute value induces the given one on K.

Finally, one proves that K_v is complete. Let $\{\xi_n\}$ be a Cauchy sequence in K_v. For each n, we can find an element $x_n \in K$ such that $|\xi_n - x_n| < 1/n$. Then one verifies immediately that $\{x_n\}$ is a Cauchy sequence in K. We let ξ be its limit in K_v. By a three-ϵ argument, one sees that $\{\xi_n\}$ converges to ξ, thus proving the completeness.

A pair (K_v, i) as in Proposition 2.1 may be called a **completion** of K. The standard pair obtained by the preceding construction could be called **the completion** of K.

Let K have a non-trivial archimedean absolute value v. If one knows that the restriction of v to the rationals is dependent on the ordinary absolute value, then the completion K_v is a complete field, containing the completion of **Q** as a closed subfield, i.e. containing the real numbers **R** as a closed subfield. It will be worthwhile to state the theorem of Gelfand-Mazur concerning the structure of such fields. First we define the notion of normed vector space.

Let K be a field with a non-trivial absolute value, and let E be a vector space over K. By a **norm** on E (compatible with the absolute value of K) we shall mean a function $\xi \to |\xi|$ of E into the real numbers such that:

NO 1. $|\xi| \geqq 0$ for all $\xi \in E$, and $= 0$ if and only if $\xi = 0$.

NO 2. For all $x \in K$ and $\xi \in E$ we have $|x\xi| = |x||\xi|$.

NO 3. If $\xi, \xi' \in E$ then $|\xi + \xi'| \leqq |\xi| + |\xi'|$.

Two norms $|\ |_1$ and $|\ |_2$ are called **equivalent** if there exist numbers $C_1, C_2 > 0$ such that for all $\xi \in E$ we have

$$C_1|\xi|_1 \leqq |\xi|_2 \leqq C_2|\xi|_1.$$

Suppose that E is finite dimensional, and let $\omega_1, \ldots, \omega_n$ be a basis of E over K. If we write an element

$$\xi = x_1\omega_1 + \cdots + x_n\omega_n$$

in terms of this basis, with $x_i \in K$, then we can define a norm by putting

$$|\xi| = \max_i |x_i|.$$

The three properties defining a norm are trivially satisfied.

Proposition 2.2. *Let K be a complete field under a non-trivial absolute value, and let E be a finite-dimensional space over K. Then any two norms on E (compatible with the given absolute value on K) are equivalent.*

Proof. We shall first prove that the topology on E is that of a product space, i.e. if $\omega_1, \ldots, \omega_n$ is a basis of E over K, then a sequence

$$\xi^{(\nu)} = x_1^{(\nu)}\omega_1 + \cdots + x_n^{(\nu)}\omega_n, \qquad x_i^{(\nu)} \in K,$$

is a Cauchy sequence in E only if each one of the n sequences $x_i^{(\nu)}$ is a Cauchy sequence in K. We do this by induction on n. It is obvious for $n = 1$. Assume $n \geqq 2$. We consider a sequence as above, and without loss of generality, we may assume that it converges to 0. (If necessary, consider $\xi^{(\nu)} - \xi^{(\mu)}$ for $\nu, \mu \to \infty$.) We must then show that the sequences of the coefficients converge to 0 also. If this is not the case, then there exists a number $a > 0$ such that we have for some j, say $j = 1$,

$$|x_1^{(\nu)}| > a$$

for arbitrarily large ν. Thus for a subsequence of (ν), $\xi^{(\nu)}/x_1^{(\nu)}$ converges to 0, and we can write

$$\frac{\xi^{(\nu)}}{x_1^{(\nu)}} - \omega_1 = \frac{x_2^{(\nu)}}{x_1^{(\nu)}}\omega_2 + \cdots + \frac{x_n^{(\nu)}}{x_1^{(\nu)}}\omega_n.$$

We let $\eta^{(\nu)}$ be the right-hand side of this equation. Then the subsequence $\eta^{(\nu)}$ converges (according to the left-hand side of our equation). By induction, we

conclude that its coefficients in terms of $\omega_2, \ldots, \omega_n$ also converge in K, say to $y_2, \ldots, y_n$. Taking the limit, we get

$$\omega_1 = y_2\omega_2 + \cdots + y_n\omega_n,$$

contradicting the linear independence of the ω_i.

We must finally see that two norms inducing the same topology are equivalent. Let $|\ |_1$ and $|\ |_2$ be these norms. There exists a number $C > 0$ such that for any $\xi \in E$ we have

$$|\xi|_1 \leqq C \quad \text{implies} \quad |\xi|_2 \leqq 1.$$

Let $a \in K$ be such that $0 < |a| < 1$. For every $\xi \in E$ there exists a unique integer s such that

$$C|a| < |a^s\xi|_1 \leqq C.$$

Hence $|a^s\xi|_2 \leqq 1$ whence we get at once

$$|\xi|_2 \leqq C^{-1}|a|^{-1}|\xi|_1.$$

The other inequality follows by symmetry, with a similar constant.

Theorem 2.3. (Gelfand-Mazur). *Let A be a commutative algebra over the real numbers, and assume that A contains an element j such that $j^2 = -1$. Let $\mathbf{C} = \mathbf{R} + \mathbf{R}j$. Assume that A is normed (as a vector space over $\mathbf{R}$), and that $|xy| \leqq |x|\,|y|$ for all $x, y \in A$. Given $x_0 \in A$, $x_0 \neq 0$, there exists an element $c \in \mathbf{C}$ such that $x_0 - c$ is not invertible in A.*

Proof. (Tornheim). Assume that $x_0 - z$ is invertible for all $z \in \mathbf{C}$. Consider the mapping $f : \mathbf{C} \to A$ defined by

$$f(z) = (x_0 - z)^{-1}.$$

It is easily verified (as usual) that taking inverses is a continuous operation. Hence f is continuous, and for $z \neq 0$ we have

$$f(z) = z^{-1}(x_0 z^{-1} - 1)^{-1} = \frac{1}{z}\left(\frac{1}{\dfrac{x_0}{z} - 1}\right).$$

From this we see that $f(z)$ approaches 0 when z goes to infinity (in $\mathbf{C}$). Hence the map $z \mapsto |f(z)|$ is a continuous map of $\mathbf{C}$ into the real numbers $\geqq 0$, is bounded, and is small outside some large circle. Hence it has a maximum, say M. Let D

be the set of elements $z \in \mathbf{C}$ such that $|f(z)| = M$. Then D is not empty; D is bounded and closed. We shall prove that D is open, hence a contradiction.

Let c_0 be a point of D, which, after a translation, we may assume to be the origin. We shall see that if r is real > 0 and small, then all points on the circle of radius r lie in D. Indeed, consider the sum

$$S(n) = \frac{1}{n} \sum_{k=1}^{n} \frac{1}{x_0 - \omega^k r}$$

where ω is a primitive n-th root of unity. Taking formally the logarithmic derivative of $X^n - r^n = \prod_{k=1}^{n} (X - \omega^k r)$ shows that

$$\frac{nX^{n-1}}{X^n - r^n} = \sum_{k=1}^{n} \frac{1}{X - \omega^k r},$$

and hence, dividing by n, and by X^{n-1}, and substituting x_0 for X, we obtain

$$S(n) = \frac{1}{x_0 - r(r/x_0)^{n-1}}.$$

If r is small (say $|r/x_0| < 1$), then we see that

$$\lim_{n \to \infty} |S(n)| = \left| \frac{1}{x_0} \right| = M.$$

Suppose that there exists a complex number λ of absolute value 1 such that

$$\left| \frac{1}{x_0 - \lambda r} \right| < M.$$

Then there exists an interval on the unit circle near λ, and there exists $\epsilon > 0$ such that for all roots of unity ζ lying in this interval, we have

$$\left| \frac{1}{x_0 - \zeta r} \right| < M - \epsilon.$$

(This is true by continuity.) Let us take n very large. Let b_n be the number of n-th roots of unity lying in our interval. Then b_n/n is approximately equal to the length of the interval (times 2π): We can express $S(n)$ as a sum

$$S(n) = \frac{1}{n} \left[\sum\nolimits_{\mathrm{I}} \frac{1}{x_0 - \omega^k r} + \sum\nolimits_{\mathrm{II}} \frac{1}{x_0 - \omega^k r} \right],$$

the first sum $\sum_{\text{I}}$ being taken over those roots of unity ω^k lying in our interval, and the second sum being taken over the others. Each term in the second sum has norm $\leqq M$ because M is a maximum. Hence we obtain the estimate

$$\begin{aligned}|S(n)| &\leqq \frac{1}{n}[|\textstyle\sum_{\text{I}}| + |\sum_{\text{II}}|] \\ &\leqq \frac{1}{n}(b_n(M - \epsilon) + (n - b_n)M) \\ &\leqq M - \frac{b_n}{n}\epsilon.\end{aligned}$$

This contradicts the fact that the limit of $|S(n)|$ is equal to M.

Corollary 2.4. *Let K be a field, which is an extension of* **R**, *and has an absolute value extending the ordinary absolute value on* **R**. *Then $K =$* **R** *or $K =$* **C**.

Proof. Assume first that K contains **C**. Then the assumption that K is a field and Theorem 2.3 imply that $K =$ **C**.

If K does not contain **C**, in other words, does not contain a square root of -1, we let $L = K(j)$ where $j^2 = -1$. We define a norm on L (as an **R**-space) by putting

$$|x + yj| = |x| + |y|$$

for $x, y \in K$. This clearly makes L into a normed **R**-space. Furthermore, if $z = x + yj$ and $z' = x' + y'j$ are in L, then

$$\begin{aligned}|zz'| &= |xx' - yy'| + |xy' + x'y| \\ &\leqq |xx'| + |yy'| + |xy'| + |x'y| \\ &\leqq |x||x'| + |y||y'| + |x||y'| + |x'||y| \\ &\leqq (|x| + |y|)(|x'| + |y'|) \\ &\leqq |z||z'|,\end{aligned}$$

and we can therefore apply Theorem 2.3 again to conclude the proof.

As an important application of Proposition 2.2, we have:

Proposition 2.5. *Let K be complete with respect to a nontrivial absolute value v. If E is any algebraic extension of K, then v has a unique extension to E. If E is finite over K, then E is complete.*

Proof. In the archimedean case, the existence is obvious since we deal with the real and complex numbers. In the non-archimedean case, we postpone

the existence proof to a later section. It uses entirely different ideas from the present ones. As to uniqueness, we may assume that E is finite over K. By Proposition 2.2, an extension of v to E defines the same topology as the max norm obtained in terms of a basis as above. Given a Cauchy sequence $\xi^{(\nu)}$ in E,

$$\xi^{(\nu)} = x_{\nu 1}\omega_1 + \cdots + x_{\nu n}\omega_n,$$

the n sequences $\{x_{\nu i}\}$ $(i = 1, \ldots, n)$ must be Cauchy sequences in K by the definition of the max norm. If $\{x_{\nu i}\}$ converges to an element z_i in K, then it is clear that the sequence $\xi^{(\nu)}$ converges to $z_1\omega_1 + \cdots + z_n\omega_n$. Hence E is complete. Furthermore, since any two extensions of v to E are equivalent, we can apply Proposition 1.1, and we see that we must have $\lambda = 1$, since the extensions induce the same absolute value v on K. This proves what we want.

From the uniqueness we can get an explicit determination of the absolute value on an algebraic extension of K. Observe first that if E is a normal extension of K, and σ is an automorphism of E over K, then the function

$$x \mapsto |\sigma x|$$

is an absolute value on E extending that of K. Hence we must have

$$|\sigma x| = |x|$$

for all $x \in E$. If E is algebraic over K, and σ is an embedding of E over K in K^a, then the same conclusion remains valid, as one sees immediately by embedding E in a normal extension of K. In particular, if α is algebraic over K, of degree n, and if $\alpha_1, \ldots, \alpha_n$ are its conjugates (counting multiplicities, equal to the degree of inseparability), then all the absolute values $|\alpha_i|$ are equal. Denoting by N the norm from $K(\alpha)$ to K, we see that

$$|N(\alpha)| = |\alpha|^n,$$

and taking the n-th root, we get:

Proposition 2.6. *Let K be complete with respect to a non-trivial absolute value. Let α be algebraic over K, and let N be the norm from $K(\alpha)$ to K. Let $n = [K(\alpha) : K]$. Then*

$$|\alpha| = |N(\alpha)|^{1/n}.$$

In the special case of the complex numbers over the real numbers, we can write $\alpha = a + bi$ with $a, b \in \mathbf{R}$, and we see that the formula of Proposition 2.6 is a generalization of the formula for the absolute value of a complex number,

$$\alpha = (a^2 + b^2)^{1/2},$$

since $a^2 + b^2$ is none other than the norm of α from $\mathbf{C}$ to $\mathbf{R}$.

Comments and examples. The process of completion is widespread in mathematics. The first example occurs in getting the real numbers from the rational numbers, with the added property of ordering. I carry this process out in full in [La 90a], Chapter IX, §3. In all other examples I know, the ordering property does not intervene. We have seen examples of completions of fields in this chapter, especially with the p-adic absolute values which are far away from ordering the field. But the real numbers are nevertheless needed as the range of values of absolute values, or more generally norms.

In analysis, one completes various spaces with various norms. Let V be a vector space over the complex numbers, say. For many applications, one must also deal with a seminorm, which satisfies the same conditions except that in **NO 1** we require only that $\|\xi\| \geqq 0$. We allow $\|\xi\| = 0$ even if $\xi \neq 0$.

One may then form the space of Cauchy sequences, the subspace of null sequences, and the factor space $\bar{V}$. The seminorm can be extended to a seminorm on $\bar{V}$ by continuity, and this extension actually turns out to be a norm. It is a general fact that $\bar{V}$ is then complete under this extension. A **Banach space** is a complete normed vector space.

Example. Let V be the vector space of step functions on $\mathbf{R}$, a step function being a complex valued function which is a finite sum of characteristic functions of intervals (closed, open, or semiclosed, i.e. the intervals may or may not contain their endpoints). For $f \in V$ we define the **$\mathbf{L^1}$-seminorm** by

$$\|f\|_1 = \int_{\mathbf{R}} |f(x)|\, dx.$$

The completion of V with respect to this seminorm is defined to be $L^1(\mathbf{R})$. One then wants to get a better idea of what elements of $L^1(\mathbf{R})$ look like. It is a simple lemma that given an L^1-Cauchy sequence in V, and given $\varepsilon > 0$, there exists a subsequence which converges uniformly except on a set of measure less than ε. Thus elements of $L^1(\mathbf{R})$ can be identified with pointwise limits of L^1-Cauchy sequences in V. The reader will find details carried out in [La 85].

Analysts use other norms or seminorms, of course, and other spaces, such as the space of C^∞ functions on $\mathbf{R}$ with compact support, and norms which may bound the derivatives. There is no end to the possible variations.

Theorem 2.3 and Corollary 2.4 are also used in the theory of Banach algebras, representing a certain type of Banach algebra as the algebra of continuous functions on a compact space, with the Gelfand-Mazur and Gelfand-Naimark theorems. Cf. [Ri 60] and [Ru 73].

Arithmetic example. For p-adic Banach spaces in connection with the number theoretic work of Dwork, see for instance Serre [Se 62], or also [La 90b], Chapter 15.

In this book we limit ourselves to complete fields and their finite extensions.

Bibliography

[La 85] S. Lang, *Real and Functional Analysis*, Springer Verlag, 1993

[La 90a] S. Lang, *Undergraduate Algebra*, Second Edition, Springer Verlag, 1990

[La 90b] S. Lang, *Cyclotomic Fields* I *and* II, Springer Verlag 1990 (combined from the first editions, 1978 and 1980)

[Ri 60] C. Rickart, *Banach Algebras*, Van Nostrand (1960), Theorems 1.7.1 and 4.2.2.

[Ru 73] W. Rudin, *Functional Analysis*, McGraw Hill (1973) Theorems 10.14 and 11.18.

[Se 62] J. P. Serre, Endomorphismes complêtement continus des espaces de Banach p-adiques, *Pub. Math. IHES* **12** (1962), pp. 69–85

§3. FINITE EXTENSIONS

Throughout this section we shall deal with a field K having a non-trivial absolute value v.

We wish to describe how this absolute value extends to finite extensions of K. If E is an extension of K and w is an absolute value on E extending v, then we shall write $w|v$.

If we let K_v be the completion, we know that v can be extended to K_v, and then uniquely to its algebraic closure K_v^a. If E is a finite extension of K, or even an algebraic one, then we can extend v to E by embedding E in K_v^a by an isomorphism over K, and taking the induced absolute value on E. We shall now prove that every extension of v can be obtained in this manner.

Proposition 3.1. *Let E be a finite extension of K. Let w be an absolute value on E extending v, and let E_w be the completion. Let K_w be the closure of K in E_w and identify E in E_w. Then $E_w = EK_w$ (the composite field).*

Proof. We observe that K_w is a completion of K, and that the composite field EK_w is algebraic over K_w and therefore complete by Proposition 2.5. Since it contains E, it follows that E is dense in it, and hence that $E_w = EK_w$.

If we start with an embedding $\sigma : E \to K_v^a$ (always assumed to be over K), then we know again by Proposition 2.5 that $\sigma E \cdot K_v$ is complete. Thus this construction and the construction of the proposition are essentially the same, up to an isomorphism. In the future, we take the embedding point of view. We must now determine when two embeddings give us the same absolute value on E.

Given two embeddings $\sigma, \tau : E \to K_v^a$, we shall say that they are **conjugate over K_v** if there exists an automorphism λ of K_v^a over K_v such that $\sigma = \lambda\tau$. We see that actually λ is determined by its effect on τE, or $\tau E \cdot K_v$.

Proposition 3.2. *Let E be an algebraic extension of K. Two embeddings $\sigma, \tau : E \to K_v^a$ give rise to the same absolute value on E if and only if they are conjugate over K_v.*

Proof. Suppose they are conjugate over K_v. Then the uniqueness of the extension of the absolute value from K_v to K_v^a guarantees that the induced absolute values on E are equal. Conversely, suppose this is the case. Let $\lambda : \tau E \to \sigma E$ be an isomorphism over K. We shall prove that λ extends to an isomorphism of $\tau E \cdot K_v$ onto $\sigma E \cdot K_v$ over K_v. Since τE is dense in $\tau E \cdot K_v$, an element $x \in \tau E \cdot K_v$ can be written

$$x = \lim \tau x_n$$

with $x_n \in E$. Since the absolute values induced by σ and τ on E coincide, it follows that the sequence $\lambda \tau x_n = \sigma x_n$ converges to an element of $\sigma E \cdot K_v$ which we denote by λx. One then verifies immediately that λx is independent of the particular sequence τx_n used, and that the map $\lambda : \tau E \cdot K_v \to \sigma E \cdot K_v$ is an isomorphism, which clearly leaves K_v fixed. This proves our proposition.

In view of the previous two propositions, if w is an extension of v to a finite extension E of K, then we may identify E_w and a composite extension EK_v of E and K_v. If $N = [E : K]$ is finite, then we shall call

$$N_w = [E_w : K_v]$$

the **local degree**.

Proposition 3.3. *Let E be a finite separable extension of K, of degree N. Then*

$$N = \sum_{w|v} N_w.$$

Proof. We can write $E = K(\alpha)$ for a single element α. Let $f(X)$ be its irreducible polynomial over K. Then over K_v, we have a decomposition

$$f(X) = f_1(X) \cdots f_r(X)$$

into irreducible factors $f_i(X)$. They all appear with multiplicity 1 according to our hypothesis of separability. The embeddings of E into K_v^a correspond to the maps of α onto the roots of the f_i. Two embeddings are conjugate if and only if they map α onto roots of the same polynomial f_i. On the other hand, it is clear that the local degree in each case is precisely the degree of f_i. This proves our proposition.

Proposition 3.4. *Let E be a finite extension of K. Then*

$$\sum_{w|v} [E_w : K_v] \leqq [E : K].$$

If E is purely inseparable over K, then there exists only one absolute value w on E extending v.

Proof. Let us first prove the second statement. If E is purely inseparable over K, and p^r is its inseparable degree, then $\alpha^{p^r} \in K$ for every α in E. Hence v has a unique extension to E. Consider now the general case of a finite extension, and let $F = E^{p^r}K$. Then F is separable over K and E is purely inseparable over F. By the preceding proposition,

$$\sum_{w|v} [F_w : K_v] = [F : K],$$

and for each w, we have $[E_w : F_w] \leqq [E : F]$. From this our inequality in the statement of the proposition is obvious.

Whenever v is an absolute value on K such that for any finite extension E of K we have $[E : K] = \sum_{w|v} [E_w : K_v]$ we shall say that v is **well behaved**. Suppose we have a tower of finite extensions, $L \supset E \supset K$. Let w range over the absolute values of E extending v, and u over those of L extending v. If $u|w$ then L_u contains E_w. Thus we have:

$$\begin{aligned} \sum_{u|v} [L_u : K_v] &= \sum_{w|v} \sum_{u|w} [L_u : E_w][E_w : K_v] \\ &= \sum_{w|v} [E_w : K_v] \sum_{u|w} [L_u : E_w] \\ &\leqq \sum_{w|v} [E_w : K_v][L : E] \\ &\leqq [E : K][L : E]. \end{aligned}$$

From this we immediately see that if v is well behaved, E finite over K, and w extends v on E, then w is well behaved (we must have an equality everywhere).

Let E be a finite extension of K. Let p^r be its inseparable degree. We recall that the norm of an element $\alpha \in K$ is given by the formula

$$N_K^E(\alpha) = \prod_\sigma \sigma\alpha^{p^r}$$

where σ ranges over all distinct isomorphisms of E over K (into a given algebraic closure).

If w is an absolute value extending v on E, then the norm from E_w to K_v will be called the **local norm**.

Replacing the above product by a sum, we get the trace, and the local trace. We abbreviate the trace by Tr.

Proposition 3.8. *Let E be a finite extension of K, and assume that v is well*

behaved. Let $\alpha \in E$. Then:

$$N_K^E(\alpha) = \prod_{w|v} N_{K_v}^{E_w}(\alpha)$$

$$\mathrm{Tr}_K^E(\alpha) = \sum_{w|v} \mathrm{Tr}_{K_v}^{E_w}(\alpha)$$

Proof. Suppose first that $E = K(\alpha)$, and let $f(X)$ be the irreducible polynomial of α over K. If we factor $f(X)$ into irreducible terms over K_v, then

$$f(X) = f_1(X) \cdots f_r(X)$$

where each $f_i(X)$ is irreducible, and the f_i are distinct because of our hypothesis that v is well behaved. The norm $N_K^E(\alpha)$ is equal to $(-1)^{\deg f}$ times the constant term of f, and similarly for each f_i. Since the constant term of f is equal to the product of the constant terms of the f_i, we get the first part of the proposition. The statement for the trace follows by looking at the penultimate coefficient of f and each f_i.

If E is not equal to $K(\alpha)$, then we simply use the transitivity of the norm and trace. We leave the details to the reader.

One can also argue directly on the embeddings. Let $\sigma_1, \ldots, \sigma_m$ be the distinct embeddings of E into K_v^{a} over K, and let p^r be the inseparable degree of E over K. The inseparable degree of $\sigma E \cdot K_v$ over K_v for any σ is at most equal to p^r. If we separate $\sigma_1, \ldots, \sigma_m$ into distinct conjugacy classes over K_v, then from our hypothesis that v is well behaved, we conclude at once that the inseparable degree of $\sigma_i E \cdot K_v$ over K_v must be equal to p^r also, for each i. Thus the formula giving the norm as a product over conjugates with multiplicity p^r breaks up into a product of factors corresponding to the conjugacy classes over K_v.

Taking into account Proposition 2.6, we have:

Proposition 3.6. *Let K have a well-behaved absolute value v. Let E be a finite extension of K, and $\alpha \in E$. Let*

$$N_w = [E_w : K_v]$$

for each absolute value w on E extending v. Then

$$\prod_{w|v} |\alpha|_w^{N_w} = |N_K^E(\alpha)|_v.$$

§4. VALUATIONS

In this section, we shall obtain, among other things, the existence theorem concerning the possibility of extending non-archimedean absolute values to algebraic extensions. We introduce first a generalization of the notion of non-archimedean absolute value.

Let Γ be a multiplicative commutative group. We shall say that an **ordering** is defined in Γ if we are given a subset S of Γ closed under multiplication such that Γ is the disjoint union of S, the unit element 1, and the set S^{-1} consisting of all inverses of elements of S.

If $\alpha, \beta \in \Gamma$ we define $\alpha < \beta$ to mean $\alpha\beta^{-1} \in S$. We have $\alpha < 1$ if and only if $\alpha \in S$. One easily verifies the following properties of the relation $<$:

1. For $\alpha, \beta \in \Gamma$ we have $\alpha < \beta$, or $\alpha = \beta$, or $\beta < \alpha$, and these possibilities are mutually exclusive.
2. $\alpha < \beta$ implies $\alpha\gamma < \beta\gamma$ for any $\gamma \in \Gamma$.
3. $\alpha < \beta$ and $\beta < \gamma$ implies $\alpha < \gamma$.

(Conversely, a relation satisfying the three properties gives rise to a subset S consisting of all elements < 1. However, we don't need this fact in the sequel.)

It is convenient to attach to an ordered group formally an extra element 0, such that $0\alpha = 0$, and $0 < \alpha$ for all $\alpha \in \Gamma$. The ordered group is then analogous to the multiplicative group of positive reals, except that there may be non-archimedean ordering.

If $\alpha \in \Gamma$ and n is an integer $\neq 0$, such that $\alpha^n = 1$, then $\alpha = 1$. This follows at once from the assumption that S is closed under multiplication and does not contain 1. In particular, the map $\alpha \mapsto \alpha^n$ is injective.

Let K be a field. By a **valuation** of K we shall mean a map $x \mapsto |x|$ of K into an ordered group Γ, together with the extra element 0, such that:

VAL 1. $|x| = 0$ if and only if $x = 0$.

VAL 2. $|xy| = |x||y|$ for all $x, y \in K$.

VAL 3. $|x + y| \leqq \max(|x|, |y|)$.

We see that a valuation gives rise to a homomorphism of the multiplicative group K^* into Γ. The valuation is called **trivial** if it maps K^* on 1. If the map giving the valuation is not surjective, then its image is an ordered subgroup of Γ, and by taking its restriction to this image, we obtain a valuation onto an ordered group, called the **value group**.

We shall denote valuations also by v. If v_1, v_2 are two valuations of K, we shall say that they are **equivalent** if there exists an order-preserving isomorphism λ of the image of v_1 onto the image of v_2 such that

$$|x|_2 = \lambda|x|_1$$

for all $x \in K$. (We agree that $\lambda(0) = 0$.)

Valuations have additional properties, like absolute values. For instance, $|1| = 1$ because $|1| = |1|^2$. Furthermore,

$$|\pm x| = |x|$$

for all $x \in K$. Proof obvious. Also, if $|x| < |y|$ then

$$|x + y| = |y|.$$

To see this, note that under our hypothesis, we have

$$|y| = |y + x - x| \leqq \max(|y + x|, |x|) = |x + y| \leqq \max(|x|, |y|) = |y|.$$

Finally, in a sum

$$x_1 + \cdots + x_n = 0,$$

at least two elements of the sum have the same value. This is an immediate consequence of the preceding remark.

Let K be a field. A subring $\mathfrak{o}$ of K is called a **valuation ring** if it has the property that for any $x \in K$ we have $x \in \mathfrak{o}$ or $x^{-1} \in \mathfrak{o}$.

We shall now see that valuation rings give rise to valuations. Let $\mathfrak{o}$ be a valuation ring of K and let U be the group of units of $\mathfrak{o}$. We contend that $\mathfrak{o}$ is a local ring. Indeed suppose that $x, y \in \mathfrak{o}$ are not units. Say $x/y \in \mathfrak{o}$. Then

$$1 + x/y = (x + y)/y \in \mathfrak{o}.$$

If $x + y$ were a unit then $1/y \in \mathfrak{o}$, contradicting the assumption that y is not a unit. Hence $x + y$ is not a unit. One sees trivially that for $z \in \mathfrak{o}$, zx is not a unit. Hence the nonunits form an ideal, which must therefore be the unique maximal ideal of $\mathfrak{o}$.

Let $\mathfrak{m}$ be the maximal ideal of $\mathfrak{o}$ and let $\mathfrak{m}^*$ be the multiplicative system of nonzero elements of $\mathfrak{m}$. Then

$$K^* = \mathfrak{m}^* \cup U \cup \mathfrak{m}^{*-1}$$

is the disjoint union of $\mathfrak{m}^*$, U, and $\mathfrak{m}^{*-1}$. The factor group K^*/U can now be given an ordering. If $x \in K^*$, we denote the coset xU by $|x|$. We put $|0| = 0$. We define $|x| < 1$ (i.e. $|x| \in S$) if and only if $x \in \mathfrak{m}^*$. Our set S is clearly closed under multiplication, and if we let $\Gamma = K^*/U$ then Γ is the disjoint union of S, 1, S^{-1}. In this way we obtain a valuation of K.

We note that if $x, y \in K$ and $x, y \neq 0$, then

$$|x| < |y| \Leftrightarrow |x/y| < 1 \Leftrightarrow x/y \in \mathfrak{m}^*.$$

Conversely, given a valuation of K into an ordered group we let $\mathfrak{o}$ be the subset of K consisting of all x such that $|x| < 1$. It follows at once from the

axioms of a valuation that $\mathfrak{o}$ is a ring. If $|x| < 1$ then $|x^{-1}| > 1$ so that x^{-1} is not in $\mathfrak{o}$. If $|x| = 1$ then $|x^{-1}| = 1$. We see that $\mathfrak{o}$ is a valuation ring, whose maximal ideal consists of those elements x with $|x| < 1$ and whose units consist of those elements x with $|x| = 1$. The reader will immediately verify that there is a bijection between valuation rings of K and equivalence classes of valuations.

The extension theorem for places and valuation rings in Chapter VII now gives us immediately the extension theorem for valuations.

Theorem 4.1. *Let K be a subfield of a field L. Then a valuation on K has an extension to a valuation on L.*

Proof. Let $\mathfrak{o}$ be the valuation ring on K corresponding to the given valuation. Let $\varphi : \mathfrak{o} \to \mathfrak{o}/\mathfrak{m}$ be the canonical homomorphism on the residue class field, and extend φ to a homomorphism of a valuation ring $\mathfrak{O}$ of L as in §3 of Chapter VII. Let $\mathfrak{M}$ be the maximal ideal of $\mathfrak{O}$. Since $\mathfrak{M} \cap \mathfrak{o}$ contains $\mathfrak{m}$ but does not contain 1, it follows that $\mathfrak{M} \cap \mathfrak{o} = \mathfrak{m}$. Let U' be the group of units of $\mathfrak{O}$. Then $U' \cap K = U$ is the group of units of $\mathfrak{o}$. Hence we have a canonical injection

$$K^*/U \to L^*/U'$$

which is immediately verified to be order-preserving. Identifying K^*/U in L^*/U' we have obtained an extension of our valuation of K to a valuation of L.

Of course, when we deal with absolute values, we require that the value group be a subgroup of the multiplicative reals. Thus we must still prove something about the nature of the value group L^*/U', whenever L is algebraic over K.

Proposition 4.2. *Let L be a finite extension of K, of degree n. Let w be a valuation of L with value group Γ'. Let Γ be the value group of K. Then $(\Gamma' : \Gamma) \leqq n$.*

Proof. Let $y_1, \dots, y_r$ be elements of L whose values represent distinct cosets of Γ in Γ'. We shall prove that the y_j are linearly independent over K. In a relation $a_1 y_1 + \cdots + a_r y_r = 0$ with $a_j \in K$, $a_j \neq 0$ two terms must have the same value, say $|a_i y_i| = |a_j y_j|$ with $i \neq j$, and hence

$$|y_i| = |a_i^{-1} a_j| |y_j|.$$

This contradicts the assumption that the values of y_i, y_j $(i \neq j)$ represent distinct cosets of Γ in Γ', and proves our proposition.

Corollary 4.3. *There exists an integer $e \geqq 1$ such that the map $\gamma \mapsto \gamma^e$ induces an injective homomorphism of Γ' into Γ.*

Proof. Take e to be the index $(\Gamma' : \Gamma)$.

Corollary 4.4. *If K is a field with a valuation v whose value group is an ordered subgroup of the ordered group of positive real numbers, and if L is an algebraic extension of K, then there exists an extension of v to L whose value group is also an ordered subgroup of the positive reals.*

Proof. We know that we can extend v to a valuation w of L with some value group Γ', and the value group Γ of v can be identified with a subgroup of $\mathbf{R}^+$. By Corollary 4.3, every element of Γ' has finite period modulo Γ. Since every element of $\mathbf{R}^+$ has a unique e-th root for every integer $e \geqq 1$, we can find in an obvious way an order-preserving embedding of Γ' into $\mathbf{R}^+$ which induces the identity on Γ. In this way we get our extension of v to an absolute value on L.

Corollary 4.5. *If L is finite over K, and if Γ is infinite cyclic, then Γ' is also infinite cyclic.*

Proof. Use Corollary 4.3 and the fact that a subgroup of a cyclic group is cyclic.

We shall now strengthen our preceding proposition to a slightly stronger one. We call $(\Gamma' : \Gamma)$ the **ramification index**.

Proposition 4.6. *Let L be a finite extension of degree n of a field K, and let $\mathfrak{O}$ be a valuation ring of L. Let $\mathfrak{M}$ be its maximal ideal, let $\mathfrak{o} = \mathfrak{O} \cap K$, and let $\mathfrak{m}$ be the maximal ideal of $\mathfrak{o}$, i.e. $\mathfrak{m} = \mathfrak{M} \cap \mathfrak{o}$. Then the residue class degree $[\mathfrak{O}/\mathfrak{M} : \mathfrak{o}/\mathfrak{m}]$ is finite. If we denote it by f, and if e is the ramification index, then $ef \leqq n$.*

Proof. Let $y_1, \ldots, y_e$ be representatives in L^* of distinct cosets of Γ'/Γ and let $z_1, \ldots, z_s$ be elements of $\mathfrak{O}$ whose residue classes mod $\mathfrak{M}$ are linearly independent over $\mathfrak{o}/\mathfrak{m}$. Consider a relation

$$\sum_{i,j} a_{ij} z_j y_i = 0$$

with $a_{ij} \in K$, not all $a_{ij} = 0$. In an inner sum

$$\sum_{j=1}^{s} a_{ij} z_j,$$

divide by the coefficient $a_{i\nu}$ having the biggest valuation. We obtain a linear combination of $z_1, \ldots, z_s$ with coefficients in $\mathfrak{o}$, and at least one coefficient equal to a unit. Since $z_1, \ldots, z_s$ are linearly independent mod $\mathfrak{M}$ over $\mathfrak{o}/\mathfrak{m}$, it follows that our linear combination is a unit. Hence

$$\left| \sum_{j=1}^{s} a_{ij} z_j \right| = |a_{i\nu}|$$

for some index ν. In the sum

$$\sum_{i=1}^{e}\left(\sum_{j=1}^{s} a_{ij}z_j\right)y_i = 0$$

viewed as a sum on i, at least two terms have the same value. This contradicts the independence of $|y_1|, \ldots, |y_e|$ mod Γ just as in the proof of Proposition 4.2.

Remark. Our proof also shows that the elements $\{z_j y_i\}$ are linearly independent over K. This will be used again later.

If w is an extension of a valuation v, then the ramification index will be denoted by $e(w|v)$ and the residue class degree will be denoted by $f(w|v)$.

Proposition 4.7. *Let K be a field with a valuation v, and let $K \subset E \subset L$ be finite extensions of K. Let w be an extension of v to E and let u be an extension of w to L. Then*

$$e(u|w)e(w|v) = e(u|v),$$
$$f(u|w)f(w|v) = f(u|v).$$

Proof. Obvious.

We can express the above proposition by saying that the ramification index and the residue class degree are multiplicative in towers.

We conclude this section by relating valuation rings in a finite extension with the integral closure.

Proposition 4.8. *Let $\mathfrak{o}$ be a valuation ring in a field K. Let L be a finite extension of K. Let $\mathfrak{O}$ be a valuation ring of L lying above $\mathfrak{o}$, and $\mathfrak{M}$ its maximal ideal. Let B be the integral closure of $\mathfrak{o}$ in L, and let $\mathfrak{B} = \mathfrak{M} \cap B$. Then $\mathfrak{O}$ is equal to the local ring $B_{\mathfrak{B}}$.*

Proof. It is clear that $B_{\mathfrak{B}}$ is contained in $\mathfrak{O}$. Conversely, let x be an element of $\mathfrak{O}$. Then x satisfies an equation with coefficients in K, not all 0, say

$$a_n x^n + \cdots + a_0 = 0, \qquad a_i \in K.$$

Suppose that a_s is the coefficient having the biggest value among the a_i for the valuation associated with the valuation ring $\mathfrak{o}$, and that it is the coefficient farthest to the left having this value. Let $b_i = a_i/a_s$. Then all $b_i \in \mathfrak{o}$ and

$$b_n, \ldots, b_{s+1} \in \mathfrak{M}.$$

Divide the equation by x^s. We get

$$(b_n x^{n-s} + \cdots + b_{s+1}x + 1) + \frac{1}{x}\left(b_{s-1} + \cdots + b_0 \frac{1}{x^{s-1}}\right) = 0.$$

Let y and z be the two quantities in parentheses in the preceding equation, so that we can write

$$-y = z/x \quad \text{and} \quad -xy = z.$$

To prove our proposition it will suffice to show that y and z lie in B and that y is not in $\mathfrak{P}$.

We use Proposition 3.5 of Chapter VII. If a valuation ring of L above contains x, then it contains y because y is a polynomial in x with coefficients in Hence such a valuation ring also contains $z = -xy$. If on the other hand the valuation ring of L above contains $1/x$, then it contains z because z is a polynomial in $1/x$ with coefficients in . Hence this valuation ring also contains y. From this we conclude by Chapter VII, Proposition 3.5, that y, z lie in B.

Furthermore, since $x \in \mathfrak{O}$, and $b_n, \ldots, b_{s+1}$ are in $\mathfrak{M}$ by construction, it follows that y cannot be in $\mathfrak{M}$, and hence cannot be in $\mathfrak{P}$. This concludes the proof.

Corollary 4.9. *Let the notation be as in the proposition. Then there is only a finite number of valuation rings of L lying above .*

Proof. This comes from the fact that there is only a finite number of maximal ideals $\mathfrak{P}$ of B lying above the maximal ideal of $\mathfrak{o}$ (Corollary of Proposition 2.1, Chapter VII).

Corollary 4.10. *Let the notation be as in the proposition. Assume in addition that L is Galois over K. If $\mathfrak{O}$ and $\mathfrak{O}'$ are two valuation rings of L lying above $\mathfrak{o}$, with maximal ideals $\mathfrak{M}$, $\mathfrak{M}'$ respectively, then there exists an automorphism σ of L over K such that $\sigma\mathfrak{O} = \mathfrak{O}'$ and $\sigma\mathfrak{M} = \mathfrak{M}'$.*

Proof. Let $\mathfrak{P} = \mathfrak{O} \cap B$ and $\mathfrak{P}' = \mathfrak{O}' \cap B$. By Proposition 2.1 of Chapter VII, we know that there exists an automorphism σ of L over K such that $\sigma\mathfrak{P} = \mathfrak{P}'$. From this our assertion is obvious.

Example. Let k be a field, and let K be a finitely generated extension of transcendence degree 1. If t is a transcendence base of K over k, then K is finite algebraic over $k(t)$. Let $\mathfrak{O}$ be a valuation ring of K containing k, and assume that $\mathfrak{O}$ is $\neq K$. Let $\mathfrak{o} = \mathfrak{O} \cap k(t)$. Then $\mathfrak{o}$ is obviously a valuation ring of $k(t)$ (the

condition about inverses is *a fortiori* satisfied), and the corresponding valuation of $k(t)$ cannot be trivial. Either t or $t^{-1} \in \mathfrak{o}$. Say $t \in \mathfrak{o}$. Then $\mathfrak{o} \cap k[t]$ cannot be the zero ideal, otherwise the canonical homomorphism $\mathfrak{o} \to \mathfrak{o}/\mathfrak{m}$ of $\mathfrak{o}$ modulo its maximal ideal would induce an isomorphism on $k[t]$ and hence an isomorphism on $k(t)$, contrary to hypothesis. Hence $\mathfrak{m} \cap k[t]$ is a prime ideal $\mathfrak{p}$, generated by an irreducible polynomial $p(t)$. The local ring $k[t]_{\mathfrak{p}}$ is obviously a valuation ring, which must be $\mathfrak{o}$ because every element of $k(t)$ has an expression of type $p^r u$ where u is a unit in $k[t]_{\mathfrak{p}}$. Thus we have determined all valuation rings of $k(t)$ containing k, and we see that the value group is cyclic. Such valuations will be called discrete and are studied in greater detail below. In view of Corollary 4.5, it follows that the valuation ring $\mathfrak{O}$ of K is also discrete.

The residue class field $\mathfrak{o}/\mathfrak{m}$ is equal to $k[t]/\mathfrak{p}$ and is therefore a finite extension of k. By Proposition 4.6, it follows that $\mathfrak{O}/\mathfrak{M}$ is finite over k (if $\mathfrak{M}$ denotes the maximal ideal of $\mathfrak{O}$).

Finally, we observe that there is only a finite number of valuation rings $\mathfrak{O}$ of K containing k such that t lies in the maximal ideal of $\mathfrak{O}$. Indeed, such a valuation ring must lie above $k[t]_{\mathfrak{p}}$ where $\mathfrak{p} = (t)$ is the prime ideal generated by t, and we can apply Corollary 4.9.

§5. COMPLETIONS AND VALUATIONS

Throughout this section, we deal with a non-archimedean absolute value v on a field K. This absolute value is then a valuation, whose value group Γ_K is a subgroup of the positive reals. We let $\mathfrak{o}$ be its valuation ring, $\mathfrak{m}$ the maximal ideal.

Let us denote by $\hat{K}$ the completion of K at v, and let $\hat{\mathfrak{o}}$ (resp. $\hat{\mathfrak{m}}$) be the closure of $\mathfrak{o}$ (resp. $\mathfrak{m}$) in $\hat{K}$. By continuity, every element of $\hat{\mathfrak{o}}$ has value $\leqq 1$, and every element of $\hat{K}$ which is not in $\hat{\mathfrak{o}}$ has value > 1. If $x \in \hat{K}$ then there exists an element $y \in K$ such that $|x - y|$ is very small, and hence $|x| = |y|$ for such an element y (by the non-archimedean property). Hence $\hat{\mathfrak{o}}$ is a valuation ring in $\hat{K}$, and $\hat{\mathfrak{m}}$ is its maximal ideal. Furthermore,

$$\hat{\mathfrak{o}} \cap K = \mathfrak{o} \quad \text{and} \quad \hat{\mathfrak{m}} \cap K = \mathfrak{m},$$

and we have an isomorphism

$$\mathfrak{o}/\mathfrak{m} \xrightarrow{\approx} \hat{\mathfrak{o}}/\hat{\mathfrak{m}}.$$

Thus the residue class field $\mathfrak{o}/\mathfrak{m}$ does not change under completion.

Let E be an extension of K, and let $\mathfrak{o}_E$ be a valuation ring of E lying above $\mathfrak{o}$. Let $\mathfrak{m}_E$ be its maximal ideal. We assume that the valuation corresponding to $\mathfrak{o}_E$ is in fact an absolute value, so that we can form the completion E. We then have

a commutative diagram:

$$\begin{array}{ccc} \mathfrak{o}_E/\mathfrak{m}_E & \xrightarrow{\approx} & \hat{\mathfrak{o}}_E/\hat{\mathfrak{m}}_E \\ \uparrow & & \uparrow \\ \mathfrak{o}/\mathfrak{m} & \xrightarrow{\approx} & \hat{\mathfrak{o}}/\hat{\mathfrak{m}} \end{array}$$

the vertical arrows being injections, and the horizontal ones being isomorphisms. Thus the residue class field extension of our valuation can be studied over the completions E of K.

We have a similar remark for the ramification index. Let $\Gamma_v(K)$ and $\Gamma_v(\hat{K})$ denote the value groups of our valuation on K and $\hat{K}$ respectively (i.e. the image of the map $x \mapsto |x|$ for $x \in K^*$ and $x \in \hat{K}^*$ respectively). We saw above that $\Gamma_v(K) = \Gamma_v(\hat{K})$; in other words, the value group is the same under completion, because of the non-archimedean property. (This is of course false in the archimedean case.) If E is again an extension of K and w is an absolute value of E extending v, then we have a commutative diagram

$$\begin{array}{ccc} \Gamma_w(E) & \xrightarrow{=} & \Gamma_w(\hat{E}) \\ \uparrow & & \uparrow \\ \Gamma_v(K) & \xrightarrow{=} & \Gamma_v(\hat{K}) \end{array}$$

from which we see that the ramification index $(\Gamma_w(E) : \Gamma_v(K))$ also does not change under completion.

§6. DISCRETE VALUATIONS

A valuation is called **discrete** if its value group is cyclic. In that case, the valuation is an absolute value (if we consider the value group as a subgroup of the positive reals). The p-adic valuation on the rational numbers is discrete for each prime number p. By Corollary 4.5, an extension of a discrete valuation to a finite extension field is also discrete. Aside from the absolute values obtained by embedding a field into the reals or complex numbers, discrete valuations are the most important ones in practice. We shall make some remarks concerning them.

Let v be a discrete valuation on a field K, and let $\mathfrak{o}$ be its valuation ring. Let $\mathfrak{m}$ be the maximal ideal. There exists an element π of $\mathfrak{m}$ which is such that its value $|\pi|$ generates the value group. (The other generator of the value group is $|\pi^{-1}|$.) Such an element π is called a **local parameter** for v (or for $\mathfrak{m}$). Every

element x of K can be written in the form

$$x = u\pi^r$$

with some unit u of $\mathfrak{o}$, and some integer r. Indeed, we have $|x| = |\pi|^r = |\pi^r|$ for some $r \in \mathbf{Z}$, whence x/π^r is a unit in $\mathfrak{o}$. We call r the **order** of x at v. It is obviously independent of the choice of parameter selected. We also say that x has a **zero of order r**. (If r is negative, we say that x has a **pole** of order $-r$.)

In particular, we see that $\mathfrak{m}$ is a principal ideal, generated by π. As an exercise, we leave it to the reader to verify that every ideal of $\mathfrak{o}$ is principal, and is a power of $\mathfrak{m}$. Furthermore, we observe that $\mathfrak{o}$ is a factorial ring with exactly one prime element (up to units), namely π.

If $x, y \in K$, we shall write $x \sim y$ if $|x| = |y|$. Let π_i $(i = 1, 2, \ldots)$ be a sequence of elements of $\mathfrak{o}$ such that $\pi_i \sim \pi^i$. Let R be a set of representatives of $\mathfrak{o}/\mathfrak{m}$ in $\mathfrak{o}$. This means that the canonical map $\mathfrak{o} \to \mathfrak{o}/\mathfrak{m}$ induces a bijection of R onto $\mathfrak{o}/\mathfrak{m}$.

Assume that K is complete under our valuation. Then every element x of $\mathfrak{o}$ can be written as a convergent series

$$x = a_0 + a_1\pi_1 + a_2\pi_2 + \cdots$$

with $a_i \in R$, and the a_i are uniquely determined by x.

This is easily proved by a recursive argument. Suppose we have written

$$x \equiv a_0 + \cdots + a_n\pi_n \pmod{\mathfrak{m}^{n+1}}$$

then $x - (a_0 + \cdots + a_n\pi_n) = \pi_{n+1}y$ for some $y \in \mathfrak{o}$. By hypothesis, we can write $y = a_{n+1} + \pi z$ with some $a_{n+1} \in R$. From this we get

$$x \equiv a_0 + \cdots + a_{n+1}\pi_{n+1} \pmod{\mathfrak{m}^{n+2}},$$

and it is clear that the n-th term in our series tends to 0. Therefore our series converges (by the non-archimedean behavior!). The fact that R contains precisely one representative of each residue class mod $\mathfrak{m}$ implies that the a_i are uniquely determined.

Examples. Consider first the case of the rational numbers with the p-adic valuation v_p. The completion is denoted by $\mathbf{Q}_p$. It is the field of ***p*-adic numbers**. The closure of $\mathbf{Z}$ in $\mathbf{Q}_p$ is the ring of ***p*-adic integers $\mathbf{Z}_p$**. We note that the prime number p is a prime element in both $\mathbf{Z}$ and its closure $\mathbf{Z}_p$. We can select our set of representatives R to be the set of integers $(0, 1, \ldots, p - 1)$. Thus every p-adic integer can be written uniquely as a convergent sum $\sum a_i p^i$ where a_i is an integer, $0 \leqq a_i \leqq p - 1$. This sum is called its p-adic expansion. Such sums are added and multiplied in the ordinary manner for convergent series.

For instance, we have the usual formalism of geometric series, and if we take $p = 3$, then

$$-1 = \frac{2}{1-3} = 2(1 + 3 + 3^2 + \cdots).$$

We note that the representatives $(0, 1, \ldots, p - 1)$ are by no means the only ones which can be used. In fact, it can be shown that $\mathbf{Z}_p$ contains the $(p - 1)$-th roots of unity, and it is often more convenient to select these roots of unity as representatives for the non-zero elements of the residue class field.

Next consider the case of a rational field $k(t)$, where k is any field and t is transcendental over k. We have a valuation determined by the prime element t in the ring $k[t]$. This valuation is discrete, and the completion of $k[t]$ under this valuation is the power series ring $k[[t]]$. In that case, we can take the elements of k itself as repersentatives of the residue class field, which is canonically isomorphic to k. The maximal ideal of $k[[t]]$ is the ideal generated by t.

This situation amounts to an algebraization of the usual situation arising in the theory of complex variables. For instance, let z_0 be a point in the complex plane. Let $\mathfrak{o}$ be the ring of functions which are holomorphic in some disc around z_0. Then $\mathfrak{o}$ is a discrete valuation ring, whose maximal ideal consists of those functions having a zero at z_0. Every element of $\mathfrak{o}$ has a power series expansion

$$f(z) = \sum_{\nu=m}^{\infty} a_\nu (z - z_0)^\nu.$$

The representatives of the residue class field can be taken to be complex numbers, a_ν. If $a_m \neq 0$, then we say that $f(z)$ has a zero of order m. The order is the same, whether viewed as order with respect to the discrete valuation in the algebraic sense, or the order in the sense of the theory of complex variables. We can select a canonical uniformizing parameter namely $z - z_0$, and

$$f(z) = (z - z_0)^m g(z)$$

where $g(z)$ is a power series beginning with a non-zero constant. Thus $g(z)$ is invertible.

Let K be again complete under a discrete valuation, and let E be a finite extension of K. Let $\mathfrak{o}_E$, $\mathfrak{m}_E$ be the valuation ring and maximal ideal in E lying above $\mathfrak{o}$, $\mathfrak{m}$ in K. Let Π be a prime element in E. If Γ_E and Γ_K are the value groups of the valuations in E and K respectively, and

$$e = (\Gamma_E : \Gamma_K)$$

is the ramification index, then

$$|\Pi^e| = |\pi|,$$

and the elements

$$\Pi^i \pi^j, \qquad 0 \leqq i \leqq e - 1, j = 0, 1, 2, \ldots$$

have order $je + i$ in E.

Let $\omega_1, \ldots, \omega_f$ be elements of E such that their residue classes mod $\mathfrak{m}_E$ from a basis of $\mathfrak{o}_E/\mathfrak{m}_E$. If R is as before a set of representatives of $\mathfrak{o}/\mathfrak{m}$ in $\mathfrak{o}$, then the set consisting of all elements

$$a_1\omega_1 + \cdots + a_f\omega_f$$

with $a_j \in R$ is a set of representatives of $\mathfrak{o}_E/\mathfrak{m}_E$ in $\mathfrak{o}_E$. From this we see that every element of $\mathfrak{o}_E$ admits a convergent expansion

$$\sum_{i=0}^{e-1} \sum_{\nu=1}^{f} \sum_{j=0}^{\infty} a_{\nu, i, j} \pi^j \omega_\nu \Pi^i.$$

Thus the elements $\{\omega_\nu \Pi^i\}$ form a set of generators of $\mathfrak{o}_E$ as a module over $\mathfrak{o}$. On the other hand, we have seen in the proof of Proposition 4.6 that these elements are linearly independent over K. Hence we obtain:

Proposition 6.1. *Let K be complete under a discrete valuation. Let E be a finite extension of K, and let e, f be the ramification index and residue class degree respectively. Then*

$$ef = [E : K].$$

Corollary 6.2. *Let $\alpha \in E$, $\alpha \neq 0$. Let v be the valuation on K and w its extension to E. Then*

$$\operatorname{ord}_v N_K^E(\alpha) = f(w \mid v) \operatorname{ord}_w \alpha.$$

Proof. This is immediate from the formula

$$|N_K^E(\alpha)| = |\alpha|^{ef}$$

and the definitions.

Corollary 6.3. *Let K be any field and v a discrete valuation on K. Let E be a finite extension of K. If v is well behaved in E (for instance if E is separable over K), then*

$$\sum_{w \mid v} e(w \mid v) f(w \mid v) = [E : K].$$

If E is Galois over K, then all e_w are equal to the same number e, all f_w are

equal to the same number f, and so

$$efr = [E:K],$$

where r is the number of extensions of v to E.

Proof. Our first assertion comes from our assumption, and Proposition 3.3. If E is Galois over K, we know from Corollary 4.10 that any two valuations of E lying above v are conjugate. Hence all ramification indices are equal, and similarly for the residue class degrees. Our relation $efr = [E:K]$ is then obvious.

§7. ZEROS OF POLYNOMIALS IN COMPLETE FIELDS

Let K be complete under a non-trivial absolute value.

Let

$$f(X) = \prod (X - \alpha_i)^{r_i}$$

be a polynomial in $K[X]$ having leading coefficient 1, and assume the roots α_i are distinct, with multiplicities r_i. Let d be the degree of f. Let g be another polynomial with coefficients in K^a, and assume that the degree of g is also d, and that g has leading coefficient 1. We let $|g|$ be the maximum of the absolute values of the coefficients of g. One sees easily that if $|g|$ is bounded, then the absolute values of the roots of g are also bounded.

Suppose that g comes close to f, in the sense that $|f - g|$ is small. If β is any root of g, then

$$|f(\beta) - g(\beta)| = |f(\beta)| = \prod |\alpha_i - \beta|^{r_i}$$

is small, and hence β must come close to some root of f. As β comes close to say $\alpha = \alpha_1$, its distance from the other roots of f approaches the distance of α_1 from the other roots, and is therefore bounded from below. In that case, we say that β **belongs to** α.

Proposition 7.1. *If g is sufficiently close to f, and $\beta_1, \ldots, \beta_s$ are the roots of g belonging to α (counting multiplicities), then $s = r_1$ is the multiplicity of α in f.*

Proof. Assume the contrary. Then we can find a sequence g_ν of polynomials approaching f with precisely s roots $\beta_1^{(\nu)}, \ldots, \beta_s^{(\nu)}$ belonging to α, but with $s \neq r$. (We can take the same multiplicity s since there is only a finite number of choices for such multiplicities.) Furthermore, the other roots of g also

belong to roots of f, and we may suppose that these roots are bunched together, according to which root of f they belong to. Since $\lim g_\nu = f$, we conclude that α must have multiplicity s in f, contradiction.

Next we investigate conditions under which a polynomial has a root in a complete field.

We assume that K is complete under a discrete valuation, with valuation ring $\mathfrak{o}$, maximal ideal $\mathfrak{p}$. We let π be a fixed prime element of $\mathfrak{p}$.

We shall deal with n-space over $\mathfrak{o}$. We denote a vector $(a_1, \ldots, a_n)$ with $a_i \in \mathfrak{o}$ by A. If $f(X_1, \ldots, X_n) \in \mathfrak{o}[X]$ is a polynomial in n variables, with integral coefficients, we shall say that A is a **zero** of f if $f(A) = 0$, and we say that A is a **zero** of f mod $\mathfrak{p}^m$ if $f(A) \equiv 0 \pmod{\mathfrak{p}^m}$.

Let $C = (c_0, \ldots, c_n)$ be in $\mathfrak{o}^{(n+1)}$. Let m be an integer $\geqq 1$. We consider the nature of the solutions of a congruence of type

$$(*) \qquad \pi^m(c_0 + c_1x_1 + \cdots + c_nx_n) \equiv 0 \pmod{\mathfrak{p}^{m+1}}.$$

This congruence is equivalent with the linear congruence

$$(**) \qquad c_0 + c_1x_1 + \cdots + c_nx_n \equiv 0 \pmod{\mathfrak{p}}.$$

If some coefficient c_i ($i = 1, \ldots, n$) is not $\equiv 0 \pmod{\mathfrak{p}}$, then the set of solutions is not empty, and has the usual structure of a solution of one inhomogeneous linear equation over the field $\mathfrak{o}/\mathfrak{p}$. In particular, it has dimension $n - 1$. A congruence (*) or (**) with some $c_i \not\equiv 0 \pmod{\mathfrak{p}}$ will be called a **proper congruence**.

As a matter of notation, we write $D_i f$ for the formal partial derivative of f with respect to X_i. We write

$$\operatorname{grad} f(X) = (D_1 f(X), \ldots, D_n f(X)).$$

Proposition 7.2. *Let $f(X) \in \mathfrak{o}[X]$. Let r be an integer $\geqq 1$ and let $A \in \mathfrak{o}^{(n)}$ be such that*

$$f(A) \equiv 0 \pmod{\mathfrak{p}^{2r-1}},$$

$$D_i f(A) \equiv 0 \pmod{\mathfrak{p}^{r-1}}, \qquad \textit{for all} \quad i = 1, \ldots, n,$$

$$D_i f(A) \not\equiv 0 \pmod{\mathfrak{p}^r}, \qquad \textit{for some } i = 1, \ldots, n.$$

Let ν be an integer $\geqq 0$ and let $B \in \mathfrak{o}^{(n)}$ be such that

$$B \equiv A \pmod{\mathfrak{p}^r} \quad \textit{and} \quad f(B) \equiv 0 \pmod{\mathfrak{p}^{2r-1+\nu}}.$$

A vector $Y \in \mathfrak{o}^{(n)}$ satisfies

$$Y \equiv B \pmod{\mathfrak{p}^{r+\nu}} \quad \textit{and} \quad f(Y) \equiv 0 \pmod{\mathfrak{p}^{2r+\nu}}$$

if and only if Y *can be written in the form* $Y = B + \pi^{r+\nu}C$, *with some* $C \in \mathfrak{o}^{(n)}$ *satisfying the proper congruence*

$$f(B) + \pi^{r+\nu} \operatorname{grad} f(B) \cdot C \equiv 0 \pmod{\mathfrak{p}^{2r+\nu}}.$$

Proof. The proof is shorter than the statement of the proposition. Write $Y = B + \pi^{r+\nu}C$. By Taylor's expansion,

$$f(B + \pi^{r+\nu}C) = f(B) + \pi^{r+\nu} \operatorname{grad} f(B) \cdot C \pmod{\mathfrak{p}^{2r+2\nu}}.$$

To solve this last congruence mod $\mathfrak{p}^{2r+\nu}$, we obtain a proper congruence by hypothesis, because $\operatorname{grad} f(B) \equiv \operatorname{grad} f(A) \equiv 0 \pmod{\mathfrak{p}^{r-1}}$.

Corollary 7.3. *Assumptions being as in Proposition* 7.2, *there exists a zero of* f *in* $\mathfrak{o}^{(n)}$ *which is congruent to* A mod $\mathfrak{p}^r$.

Proof. We can write this zero as a convergent sum

$$A + \pi^{r+1}C_1 + \pi^{r+2}C_2 + \cdots$$

solving for $C_1, C_2, \ldots$ inductively as in the proposition.

Corollary 7.4. *Let* f *be a polynomial in one variable in* $\mathfrak{o}[X]$, *and let* $a \in \mathfrak{o}$ *be such that* $f(a) \equiv 0 \pmod{\mathfrak{p}}$ *but* $f'(a) \not\equiv 0 \pmod{\mathfrak{p}}$. *Then there exists* $b \in \mathfrak{o}$, $b \equiv a \pmod{\mathfrak{p}}$ *such that* $f(b) = 0$.

Proof. Take $n = 1$ and $r = 1$ in the proposition, and apply Corollary 7.3.

Corollary 7.5. *Let* m *be a positive integer not divisible by the characteristic of* K. *There exists an integer* r *such that for any* $a \in \mathfrak{o}$, $a \equiv 1 \pmod{\mathfrak{p}^r}$, *the equation* $X^m - a = 0$ *has a root in* K.

Proof. Apply the proposition.

Example. In the 2-adic field $\mathbf{Q}_2$, there exists a square root of -7, i.e. $\sqrt{-7} \in \mathbf{Q}_2$, because $-7 = 1 - 8$.

When the absolute value is not discrete, it is still possible to formulate a criterion for a polynomial to have a zero by **Newton approximation**. (Cf. my paper, "On quasi-algebraic closure," *Annals of Math.* (1952) pp. 373–390.

Proposition 7.6. *Let* K *be a complete under a non-archimedean absolute value (nontrivial). Let* $\mathfrak{o}$ *be the valuation ring and let* $f(X) \in \mathfrak{o}[X]$ *be a polynomial in one variable. Let* $\alpha_0 \in \mathfrak{o}$ *be such that*

$$|f(\alpha_0)| < |f'(\alpha_0)^2|$$

(here f' *denotes the formal derivative of* f*). Then the sequence*

$$\alpha_{i+1} = \alpha_i - \frac{f(\alpha_i)}{f'(\alpha_i)}$$

converges to a root α of f in $\mathfrak{o}$, and we have

$$|\alpha - \alpha_0| \leqq \left|\frac{f(\alpha_0)}{f'(\alpha_0)^2}\right| < 1.$$

Proof. Let $c = |f(\alpha_0)/f'(\alpha_0)^2| < 1$. We show inductively that:

1. $|\alpha_i| \leqq 1$,

2. $|\alpha_i - \alpha_0| \leqq c$,

3. $\left|\dfrac{f(\alpha_i)}{f'(\alpha_i)^2}\right| \leqq c^{2^i}$.

These three conditions obviously imply our proposition. If $i = 0$, they are hypotheses. By induction, assume them for i. Then:

1. $|f(\alpha_i)/f'(\alpha_i)^2| \leqq c^{2^i}$ gives $|\alpha_{i+1} - \alpha_i| \leqq c^{2^i} < 1$, whence $|\alpha_{i+1}| \leqq 1$.

2. $|\alpha_{i+1} - \alpha_0| \leqq \max\{|\alpha_{i+1} - \alpha_i|, |\alpha_i - \alpha_0|\} = c$.

3. By Taylor's expansion, we have

$$f(\alpha_{i+1}) = f(\alpha_i) - f'(\alpha_i)\frac{f(\alpha_i)}{f'(\alpha_i)} + \beta\left(\frac{f(\alpha_i)}{f'(\alpha_i)}\right)^2$$

for some $\beta \in \mathfrak{o}$, and this is less than or equal to

$$\left|\frac{f(\alpha_i)}{f'(\alpha_i)}\right|^2$$

in absolute value.

Using Taylor's expansion on $f'(\alpha_{i+1})$ we conclude that

$$|f'(\alpha_{i+1})| = |f'(\alpha_i)|.$$

From this we get

$$\left|\frac{f(\alpha_{i+1})}{f'(\alpha_{i+1})^2}\right| \leqq c^{2^{i+1}}$$

as desired.

The technique of the proposition is also useful when dealing with rings, say a local ring $\mathfrak{o}$ with maximal ideal $\mathfrak{m}$ such that $\mathfrak{m}^r = 0$ for some integer $r > 0$. If one has a polynomial f in $\mathfrak{o}[X]$ and an approximate root α_0 such that

$$f'(\alpha_0) \not\equiv 0 \bmod \mathfrak{m},$$

then the Newton approximation sequence shows how to refine α_0 to a root of f.

Example in several variables. *Let K be complete under a non-archimedean absolute value. Let $f(X_1, \ldots, X_{n+1}) \in K[X]$ be a polynomial with coefficients in K. Let $(a_1, \ldots, a_n, b) \in K^{n+1}$. Assume that $f(a, b) = 0$. Let D_{n+1} be the*

partial derivative with respect to the $(n + 1)$*-th variable, and assume that* $D_{n+1}f(a, b) \neq 0$. *Let* $(\bar{a}) \in K^n$ *be sufficiently close to* (a). *Then there exists an element* $\bar{b}$ *of* K *close to* b *such that* $f(\bar{a}, \bar{b}) = 0$.

This statement is an immediate corollary of Proposition 7.6. By multiplying all a_i, b by a suitable non-zero element of K one can change them to elements of $\mathfrak{o}$. Changing the variables accordingly, one may assume without loss of generality that $a_i, b \in \mathfrak{o}$, and the condition on the partial derivative not vanishing is preserved. Hence Proposition 7.6 may be applied. After perturbing (a) to $(\bar{a})$, the element b becomes an approximate solution of $f(\bar{a}, X)$. As $(\bar{a})$ approaches (a), $f(\bar{a}, b)$ approaches 0 and $D_{n+1}f(\bar{a}, b)$ approaches $D_{n+1}f(a, b) \neq 0$. Hence for $(\bar{a})$ sufficiently close to (a), the conditions of Proposition 7.6 are satisfied, and one may refine b to a root of $f(\bar{a}, X)$, thus proving the assertion.

The result was used in a key way in my paper "On Quasi Algebraic Closure". It is the analogue of Theorem 3.6 of Chapter XI, for real fields.

In the language of algebraic geometry (which we now assume), the result can be reformulated as follows. Let V be a variety defined over K. Let P be a simple point of V in K. Then there is a whole neighborhood of simple points of V in K. Especially, suppose that V is defined by a finite number of polynomial equations over a finitely generated field k over the prime field. After a suitable projection, one may assume that the variety is affine, and defined by one equation $f(X_1, \ldots, X_{n+1}) = 0$ as in the above statement, and that the point is $P = (a_1, \ldots, a_n, b)$ as above. One can then select $\bar{a}_i = x_i$ close to a_i but such that $(x_1, \ldots, x_n)$ are algebraically independent over k. Let y be the refinement of b such that $f(x, y) = 0$. Then (x, y) is a generic point of V over k, and the coordinates of (x, y) lie in K. In geometric terms, this means that the function field of the variety can be embedded in K over k, just as Theorem 3.6 of Chapter XI gave the similar result for an embedding in a real closed field, e.g. the real numbers.

EXERCISES

1. (a) Let K be a field with a valuation. If

$$f(X) = a_0 + a_1X + \cdots + a_nX^n$$

is a polynomial in $K[X]$, define $|f|$ to be the max on the values $|a_i|$ $(i = 0, \ldots, n)$. Show that this defines an extension of the valuation to $K[X]$, and also that the valuation can be extended to the rational field $K(X)$. How is Gauss' lemma a special case of the above statement? Generalize to polynomials in several variables.

(b) Let f be a polynomial with complex coefficients. Define $|f|$ to be the maximum of the absolute values of the coefficients. Let d be an integer $\geqq 1$. Show that

there exist constants C_1, C_2 (depending only on d) such that, if f, g are polynomials in $\mathbf{C}[X]$ of degrees $\leqq d$, then

$$C_1 |f||g| \leqq |fg| \leqq C_2 |f||g|.$$

[*Hint*: Induction on the number of factors of degree 1. Note that the right inequality is trivial.]

2. Let $M_{\mathbf{Q}}$ be the set of absolute values consisting of the ordinary absolute value and all p-adic absolute values v_p on the field of rational numbers $\mathbf{Q}$. Show that for any rational number $a \in \mathbf{Q}$, $a \neq 0$, we have

$$\prod_{v \in M_{\mathbf{Q}}} |a|_v = 1.$$

If K is a finite extension of $\mathbf{Q}$, and M_K denotes the set of absolute values on K extending those of $M_{\mathbf{Q}}$, and for each $w \in M_K$ we let N_w be the local degree $[K_w : \mathbf{Q}_v]$, show that for $\alpha \in K$, $\alpha \neq 0$, we have

$$\prod_{w \in M_K} |\alpha|_w^{N_w} = 1.$$

3. Show that the p-adic numbers $\mathbf{Q}_p$ have no automorphisms other than the identity. [*Hint*: Show that such automorphisms are continuous for the p-adic topology. Use Corollary 7.5 as an algebraic characterization of elements close to 1.]

4. Let A be a principal entire ring, and let K be its quotient field. Let $\mathfrak{o}$ be a valuation ring of K containing A, and assume $\mathfrak{o} \neq K$. Show that $\mathfrak{o}$ is the local ring $A_{(p)}$ for some prime element p. [This applies both to the ring $\mathbf{Z}$ and to a polynomial ring $k[X]$ over a field k.]

5. Let A be an entire ring, and let K be its quotient field. Assume that every finitely generated ideal of A is principal. Let $\mathfrak{o}$ be a discrete valuation ring of K containing A. Show that $\mathfrak{o} = A_{(p)}$ for some element p of A, and that p is a generator of the maximal ideal of $\mathfrak{o}$.

6. Let $\mathbf{Q}_p$ be a p-adic field. Show that $\mathbf{Q}_p$ contains infinitely many quadratic fields of type $\mathbf{Q}(\sqrt{-m})$, where m is a positive integer.

7. Show that the ring of p-adic integers $\mathbf{Z}_p$ is compact. Show that the group of units in $\mathbf{Z}_p$ is compact.

8. If K is a field complete with respect to a discrete valuation, with finite residue class field, and if $\mathfrak{o}$ is the ring of elements of K whose orders are $\geqq 0$, show that $\mathfrak{o}$ is compact. Show that the group of units of $\mathfrak{o}$ is closed in $\mathfrak{o}$ and is compact.

9. Let K be a field complete with respect to a discrete valuation, let $\mathfrak{o}$ be the ring of integers of K, and assume that $\mathfrak{o}$ is compact. Let $f_1, f_2, \ldots$ be a sequence of polynomials in n variables, with coefficients in $\mathfrak{o}$. Assume that all these polynomials have degree $\leqq d$, and that they converge to a polynomial f (i.e. that $|f - f_i| \to 0$ as $i \to \infty$). If each f_i has a zero in $\mathfrak{o}$, show that f has a zero in $\mathfrak{o}$. If the polynomials f_i are homogeneous of degree d, and if each f_i has a non-trivial zero in $\mathfrak{o}$, show that f has a non-trivial zero in $\mathfrak{o}$. [*Hint*: Use the compactness of $\mathfrak{o}$ and of the units of $\mathfrak{o}$ for the homogeneous case.]

(For applications of this exercise, and also of Proposition 7.6, cf. my paper "On quasi-algebraic closure," *Annals of Math.*, **55** (1952), pp. 412–444.)

10. Show that if p, p' are two distinct prime numbers, then the fields $\mathbf{Q}_p$ and $\mathbf{Q}_{p'}$ are not isomorphic.
11. Prove that the field $\mathbf{Q}_p$ contains all $(p-1)$-th roots of unity. [*Hint*: Use Proposition 7.6, applied to the polynomial $X^{p-1} - 1$ which splits into factors of degree 1 in the residue class field.] Show that two distinct $(p-1)$-th roots of unity cannot be congruent mod p.
12. (a) Let $f(X)$ be a polynomial of degree 1 in $\mathbf{Z}[X]$. Show that the values $f(a)$ for $a \in \mathbf{Z}$ are divisible by infinitely many primes.
    (b) Let F be a finite extension of $\mathbf{Q}$. Show that there are infinitely many primes p such that all conjugates of F (in an algebraic closure of $\mathbf{Q}_p$) actually are contained in $\mathbf{Q}_p$. [*Hint*: Use the irreducible polynomial of a generator for a Galois extension of $\mathbf{Q}$ containing F.]
13. Let K be a field of characteristic 0, complete with respect to a non-archimedean absolute value. Show that the series

$$\exp(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$$

$$\log(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots$$

converge in some neighborhood of 0. (The main problem arises when the characteristic of the residue class field is $p > 0$, so that p divides the denominators $n!$ and n. Get an expression which determines the power of p occurring in $n!$.) Prove that the exp and log give mappings inverse to each other, from a neighborhood of 0 to a neighborhood of 1.

14. Let K be as in the preceding exercise, of characteristic 0, complete with respect to a non-archimedean absolute value. For every integer $n > 0$, show that the usual binomial expansion for $(1 + x)^{1/n}$ converges in some neighborhood of 0. Do this first assuming that the characteristic of the residue class field does not divide n, in which case the assertion is much simpler to prove.
15. Let F be a complete field with respect to a discrete valuation, let $\mathfrak{o}$ be the valuation ring, π a prime element, and assume that $\mathfrak{o}/(\pi) = k$. Prove that if $a, b \in \mathfrak{o}$ and $a \equiv b \pmod{\pi^r}$ with $r > 0$ then $a^{p^n} \equiv b^{p^n} \pmod{\pi^{r+n}}$ for all integers $n \geqq 0$.
16. Let F be as above. Show that there exists a system of representatives R for $\mathfrak{o}/(\pi)$ in $\mathfrak{o}$ such that $R^p = R$ and that this system is unique (Teichmüller). [*Hint*: Let α be a residue class in k. For each $\nu \geqq 0$ let a_ν be a representative in $\mathfrak{o}$ of $a^{p^{-\nu}}$ and show that the sequence $a_\nu^{p^\nu}$ converges for $\nu \to \infty$, and in fact converges to a representative a of α, independent of the choices of a_ν.] Show that the system of representatives R thus obtained is closed under multiplication, and that if F has characteristic p, then R is closed under addition, and is isomorphic to k.
17. (a) (**Witt vectors again**). Let k be a perfect field of characteristic p. We use the Witt vectors as described in the exercises of Chapter VI. One can define an absolute value on $W(k)$, namely $|x| = p^{-r}$ if x_r is the first non-zero component of x. Show that this is an absolute value, obviously discrete, defined on the ring, and which can be extended at once to the quotient field. Show that this quotient field is complete, and note that $W(k)$ is the valuation ring. The maximal ideal consists of those x such that $x_0 = 0$, i.e. is equal to $pW(k)$.

(b) Assume that F has characteristic 0. Map each vector $x \in W(k)$ on the element

$$\sum \xi_i^{p^{-i}} p^i$$

where ξ_i is a representative of x_i in the special system of Exercise 15. Show that this map is an embedding of $W(k)$ into $\mathfrak{o}$.

18. **(Local uniformization).** Let k be a field, K a finitely generated extension of transcendence degree 1, and $\mathfrak{o}$ a discrete valuation ring of K over k, with maximal ideal $\mathfrak{m}$. Assume that $\mathfrak{o}/\mathfrak{m} = k$. Let x be a generator of $\mathfrak{m}$, and assume that K is separable over $k(x)$. Show that there exists an element $y \in \mathfrak{o}$ such that $K = k(x, y)$, and also having the following property. Let φ be the place on K determined by $\mathfrak{o}$. Let $a = \varphi(x)$, $b = \varphi(y)$ (of course $a = 0$). Let $f(X, Y)$ be the irreducible polynomial in $k[X, Y]$ such that $f(x, y) = 0$. Then $D_2 f(a, b) \neq 0$. [*Hint:* Write first $K = k(x, z)$ where z is integral over $k[x]$. Let $z = z_1, \dots, z_n (n \geqq 2)$ be the conjugates of z over $k(x)$, and extend $\mathfrak{o}$ to a valuation ring $\mathfrak{O}$ of $k(x, z_1, \dots, z_n)$. Let

$$z = a_0 + a_1 x + \cdots + a_r x^r + \cdots$$

be the power series expansion of z with $a_i \in k$, and let $P_r(x) = a_0 + \cdots + a_r x^r$. For $i = 1, \dots, n$ let

$$y_i = \frac{z_i - P_r(x)}{x^r}.$$

Taking r large enough, show that y_1 has no pole at $\mathfrak{O}$ but $y_2, \dots, y_n$ have poles at $\mathfrak{O}$. The elements $y_1, \dots, y_n$ are conjugate over $k(x)$. Let $f(X, Y)$ be the irreducible polynomial of (x, y) over k. Then $f(x, Y) = \psi_n(x) Y^n + \cdots + \psi_0(x)$ with $\psi_i(x) k[x]$. We may also assume $\psi_i(0) \neq 0$ (since f is irreducible). Write $f(x, Y)$ in the form

$$f(x, Y) = \psi_n(x) y_2 \cdots y_n (Y - y_1)(y_2^{-1} Y - 1) \cdots (y_n^{-1} Y - 1).$$

Show that $\psi_n(x) y_2 \cdots y_n = u$ does not have a pole at $\mathfrak{O}$. If $w \in \mathfrak{O}$, let w denote its residue class modulo the maximal ideal of $\mathfrak{O}$. Then

$$0 \neq f(\bar{x}, Y) = (-1)^{n-1} \bar{u} (Y - \bar{y}_1).$$

Let $y = y_1$, $\bar{y} = b$. We find that $D_2 f(a, b) = (-1)^{n-1} \bar{u} \neq 0$.]

19. Prove the converse of Exercise 17, i.e. if $K = k(x, y)$, $f(X, Y)$ is the irreducible polynomial of (x, y) over k, and if $a, b \in k$ are such that $f(a, b) = 0$, but $D_2 f(a, b) \neq 0$, then there exists a unique valuation ring $\mathfrak{o}$ of K with maximal ideal $\mathfrak{m}$ such that $x \equiv a$ and $y \equiv b \pmod{\mathfrak{m}}$. Furthermore, $\mathfrak{o}/\mathfrak{m} = k$, and $x - a$ is a generator of $\mathfrak{m}$. [*Hint*: If $g(x, y) \in k[x, y]$ is such that $g(a, b) = 0$, show that $g(x, y) = (x - a) A(x, y)/B(x, y)$ where A, B are polynomials such that $B(a, b) \neq 0$. If $A(a, b) = 0$ repeat the process. Show that the process cannot be repeated indefinitely, and leads to a proof of the desired assertion.]

20. **(Iss'sa-Hironaka** *Ann. of Math* 83 (1966), pp. 34–46). This exercise requires a good working knowledge of complex variables. Let K be the field of meromorphic functions on the complex plane $\mathbf{C}$. Let $\mathfrak{O}$ be a discrete valuation ring of K (containing the

constants $\mathbf{C}$). Show that the function z is in $\mathfrak{O}$. [*Hint:* Let $a_1, a_2, \ldots$ be a discrete sequence of complex numbers tending to infinity, for instance the positive integers. Let $v_1, v_2, \ldots,$ be a sequence of integers, $0 \leqq v_i \leqq p - 1$, for some prime number p, such that $\sum \nu_i p^i$ is not the p-adic expansion of a rational number. Let f be an entire function having a zero of order $\nu_i p^i$ at a_i for each i and no other zero. If z is not in $\mathfrak{o}$, consider the quotient

$$g(z) = \frac{f(z)}{\prod_{i=1}^{n} (z - a_i)^{\nu_i p^i}}.$$

From the Weierstrass factorization of an entire function, show that $g(z) = h(z)^{p^{n+1}}$ for some entire function $h(z)$. Now analyze the zero of g at the discrete valuation of $\mathfrak{o}$ in terms of that of f and $\prod (z - a_i)^{\nu_i p^i}$ to get a contradiction.]

If U is a non-compact Riemann surface, and L is the field of meromorphic functions on U, and if $\mathfrak{o}$ is a discrete valuation ring of L containing the constants, show that every holomorphic function φ on U lies in $\mathfrak{o}$. [*Hint*: Map $\varphi : U \to \mathbf{C}$, and get a discrete valuation of K by composing φ with meromorphic functions on $\mathbf{C}$. Apply the first part of the exercise.] Show that the valuation ring is the one associated with a complex number. [*Further hint*: If you don't know about Riemann surfaces, do it for the complex plane. For each $z \in U$, let f_z be a function holomorphic on U and having only a zero of order 1 at z. If for some z_0 the function f_{z_0} has order $\geqq 1$ at $\mathfrak{o}$, then show that $\mathfrak{o}$ is the valuation ring associated with z_0. Otherwise, every function f_z has order 0 at $\mathfrak{o}$. Conclude that the valuation of $\mathfrak{o}$ is trivial on any holomorphic function by a limit trick analogous to that of the first part of the exercise.]

Part Three

LINEAR ALGEBRA and REPRESENTATIONS

We shall be concerned with modules and vector spaces, going into their structure under various points of view. The main theme here is to study a pair, consisting of a module, and an endomorphism, or a ring of endomorphisms, and try to decompose this pair into a direct sum of components whose structure can then be described explicitly. The direct sum theme recurs in every chapter. Sometimes, we use a duality to obtain our direct sum decomposition relative to a pairing, and sometimes we get our decomposition directly. If a module refuses to decompose into a direct sum of simple components, then there is no choice but to apply the Grothendieck construction and see what can be obtained from it.

The extension theme occurs only once, in Witt's theorem, in a brief counterpoint to the decomposition theme.

CHAPTER XIII

Matrices and Linear Maps

Presumably readers of this chapter will have had some basic acquaintance with linear algebra in elementary courses. We go beyond such courses by pointing out that a lot of results hold for free modules over a commutative ring. This is useful when one wants to deal with families of linear maps, and reduction modulo an ideal.

Note that §8 and §9 give examples of group theory in the context of linear groups.

Throughout this chapter, we let R be a commutative ring, and we let E, F be R-modules. We suppress the prefix R in front of linear maps and modules.

§1. MATRICES

By an $m \times n$ **matrix** in R one means a doubly indexed family of elements of R, (a_{ij}), $(i = 1, \ldots, m$ and $j = 1, \ldots, n)$, usually written in the form

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ & \cdots & \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}.$$

We call the elements a_{ij} the **coefficients** or **components** of the **matrix**. A $1 \times n$ matrix is called a **row vector** (of dimension, or size, n) and a $m \times 1$ matrix is called a **column vector** (of dimension, or size, m). In general, we say that (m, n) is the **size** of the matrix, or also $m \times n$.

We define addition for matrices of the same size by components. If $A = (a_{ij})$ and $B = (b_{ij})$ are matrices of the same size, we define $A + B$ to be the matrix whose ij-component is $a_{ij} + b_{ij}$. Addition is obviously associative. We define the multiplication of a matrix A by an element $c \in R$ to be the matrix (ca_{ij}),

whose ij-component is ca_{ij}. Then the set of $m \times n$ matrices in R is a module (i.e. an R-module).

We define the product AB of two matrices only under certain conditions. Namely, when A has size (m, n) and B has size (n, r), i.e. only when the size of the rows of A is the same as the size of the columns of B. If that is the case, let $A = (a_{ij})$ and let $B = (b_{jk})$. We define AB to be the $m \times r$ matrix whose ik-component is

$$\sum_{j=1}^{n} a_{ij} b_{jk}.$$

If A, B, C are matrices such that AB is defined and BC is defined, then so is $(AB)C$ and $A(BC)$ and we have

$$(AB)C = A(BC).$$

This is trivial to prove. If $C = (c_{kl})$, then the reader will see at once that the il-component of either of the above products is equal to

$$\sum_j \sum_k a_{ij} b_{jk} c_{kl}.$$

An $m \times n$ matrix is said to be a **square matrix** if $m = n$. For example, a 1×1 matrix is a square matrix, and will sometimes be identified with the element of R occurring as its single component.

For a given integer $n \geqq 1$ the set of square $n \times n$ matrices forms a ring.

This is again trivially verified and will be left to the reader.

The unit element of the ring of $n \times n$ matrices is the matrix

$$I_n = \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & & & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & & & & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

whose components are equal to 0 except on the diagonal, in which case they are equal to 1. We sometimes write I instead of I_n.

If $A = (a_{ij})$ is a square matrix, we define in general its **diagonal components** to be the elements a_{ii}.

We have a natural ring-homomorphism of R into the ring of $n \times n$ matrices, given by

$$c \mapsto cI_n.$$

Thus cI_n is the square $n \times n$ matrix having all its components equal to 0 except the diagonal components, which are equal to c. Let us denote the ring of $n \times n$

matrices in R by $\mathrm{Mat}_n(R)$. Then $\mathrm{Mat}_n(R)$ is an algebra over R (with respect to the above homomorphism).

Let $A = (a_{ij})$ be an $m \times n$ matrix. We define its **transpose** ${}^tA$ to be the matrix (a_{ji}) ($j = 1, \ldots, n$ and $i = 1, \ldots, m$). Then ${}^tA$ is an $n \times m$ matrix. The reader will verify at once that if A, B are of the same size, then

$${}^t(A + B) = {}^tA + {}^tB.$$

If $c \in R$ then ${}^t(cA) = c\,{}^tA$. If A, B can be multiplied, then ${}^tB\,{}^tA$ is defined and we have

$${}^t(AB) = {}^tB\,{}^tA.$$

We note the operations on matrices commute with homomorphisms. More precisely, let $\varphi : R \to R'$ be a ring-homomorphism. If A, B are matrices in R, we define φA to be the matrix obtained by applying φ to all the components of A. Then

$$\varphi(A + B) = \varphi A + \varphi B, \qquad \varphi(AB) = (\varphi A)(\varphi B), \qquad \varphi(cA) = \varphi(c)\varphi A,$$
$$\varphi({}^tA) = {}^t\varphi(A).$$

A similar remark will hold throughout our discussion of matrices (for instance in the next section).

Let $A = (a_{ij})$ be a square $n \times n$ matrix in a commutative ring R. We define the **trace** of A to be

$$\mathrm{tr}(A) = \sum_{i=1}^{n} a_{ii};$$

in other words, the trace is the sum of the diagonal elements.

If A, B are $n \times n$ matrices, then

$$\mathrm{tr}(AB) = \mathrm{tr}(BA).$$

Indeed, if $A = (a_{ij})$ and $B = (b_{ij})$ then

$$\mathrm{tr}(AB) = \sum_i \sum_\nu a_{i\nu} b_{\nu i} = \mathrm{tr}(BA).$$

As an application, we observe that if B is an invertible $n \times n$ matrix, then

$$\mathrm{tr}(B^{-1}AB) = \mathrm{tr}(A).$$

Indeed, $\mathrm{tr}(B^{-1}AB) = \mathrm{tr}(ABB^{-1}) = \mathrm{tr}(A)$.

§2. THE RANK OF A MATRIX

Let k be a field and let A be an $m \times n$ matrix in k. By the **row rank** of A we shall mean the maximum number of linearly independent rows of A, and by the **column rank** of A we shall mean the maximum number of linearly independent columns of A. Thus these ranks are the dimensions of the vector spaces generated respectively by the rows of A and the columns of A. We contend that these ranks are equal to the same number, and we define the **rank** of A to be that number.

Let $A^1, \dots, A^n$ be the columns of A, and let $A_1, \dots, A_m$ be the rows of A. Let ${}^tX = (x_1, \dots, x_m)$ have components $x_i \in k$. We have a linear map

$$X \mapsto x_1 A_1 + \cdots + x_m A_m$$

of $k^{(m)}$ onto the space generated by the row vectors. Let W be its kernel. Then W is a subspace of $k^{(m)}$ and

$$\dim W + \text{row rank} = m.$$

If Y is a column vector of dimension m, then the map

$$(X, Y) \mapsto {}^tXY = X \cdot Y$$

is a bilinear map into k, if we view the 1×1 matrix ${}^tXY$ as an element of k. We observe that W is the orthogonal space to the column vectors $A^1, \dots, A^n$, i.e. it is the space of all X such that $X \cdot A^j = 0$ for all $j = 1, \dots, n$. By the duality theorem of Chapter III, we know that $k^{(m)}$ is its own dual under the pairing

$$(X, Y) \mapsto X \cdot Y$$

and that $k^{(m)}/W$ is dual to the space generated by $A^1, \dots, A^n$. Hence

$$\dim k^{(m)}/W = \text{column rank},$$

or

$$\dim W + \text{column rank} = m.$$

From this we conclude that

$$\text{column rank} = \text{row rank},$$

as desired.

We note that W may be viewed as the space of solutions of the system of n linear equations

$$x_1 A_1 + \cdots + x_m A_m = 0,$$

in m unknowns $x_1, \dots, x_m$. Indeed, if we write out the preceding vector equation in terms of all the coordinates, we get the usual system of n linear equations. We let the reader do this if he or she wishes.

§3. MATRICES AND LINEAR MAPS

Let E be a module, and assume that there exists a basis $\mathcal{B} = \{\xi_1, \dots, \xi_n\}$ for E over R. This means that every element of E has a unique expression as a linear combination

$$x = x_1\xi_1 + \cdots + x_n\xi_n$$

with $x_i \in R$. We call $(x_1, \dots, x_n)$ the **components** of x with respect to the basis. We may view this n-tuple as a row vector. We shall denote by X the transpose of the row vector $(x_1, \dots, x_n)$. We call X the **column vector of x with respect to the basis.**

We observe that if $\{\xi'_1, \dots, \xi'_m\}$ is another basis of E over R, then $m = n$. Indeed, let $\mathfrak{p}$ be a maximal ideal of R. Then $E/\mathfrak{p}E$ is a vector space over the field $R/\mathfrak{p}R$, and it is immediately clear that if we denote by $\bar{\xi}_i$ the residue class of ξ_i mod $\mathfrak{p}E$, then $\{\bar{\xi}_1, \dots, \bar{\xi}_n\}$ is a basis for $E/\mathfrak{p}E$ over $R/\mathfrak{p}R$. Hence n is also the dimension of this vector space, and we know the invariance of the cardinality for bases of vector spaces over fields. Thus $m = n$. We shall call n the **dimension** of the module E over R.

We shall view $R^{(n)}$ as the module of column vectors of size n. It is a free module of dimension n over R. It has a basis consisting of the unit vectors $e^1, \dots, e^n$ such that

$${}^t e^i = (0, \dots, 0, 1, 0, \dots, 0)$$

has components 0 except for its i-th component, which is equal to 1.

An $m \times n$ matrix A gives rise to a linear map

$$L_A : R^{(n)} \to R^{(m)}$$

by the rule

$$X \mapsto AX.$$

Namely, we have $A(X + Y) = AX + AY$ and $A(cX) = cAX$ for column vectors X, Y and $c \in R$.

The above considerations can be extended to a slightly more general context, which can be very useful. Let E be an abelian group and assume that R is a commutative subring of

$$\mathrm{End}_{\mathbf{Z}}(E) = \mathrm{Hom}_{\mathbf{Z}}(E, E).$$

Then E is an R-module. Furthermore, if A is an $m \times n$ matrix in R, then we get a linear map

$$L_A : E^{(n)} \to E^{(m)}$$

defined by a rule similar to the above, namely $X \mapsto AX$. However, this has to be interpreted in the obvious way. If $A = (a_{ij})$ and X is a column vector of elements of E, then

$$AX = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ & \cdots & \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix},$$

where $y_i = \sum_{j=1}^{n} a_{ij} x_j$.

If A, B are matrices in R whose product is defined, then for any $c \in R$ we have

$$L_{AB} = L_A L_B \quad \text{and} \quad L_{cA} = cL_A.$$

Thus we have associativity, namely

$$A(BX) = (AB)X.$$

An arbitrary commutative ring R may be viewed as a module over itself. In this way we recover the special case of our map from $R^{(n)}$ into $R^{(m)}$. Furthermore, if E is a module over R, then R may be viewed as a ring of endomorphisms of E.

Proposition 3.1. *Let E be a free module over R, and let $\{x_1, \ldots, x_n\}$ be a basis. Let $y_1, \ldots, y_n$ be elements of E. Let A be the matrix in R such that*

$$A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

Then $\{y_1, \ldots, y_n\}$ is a basis of E if and only if A is invertible.

Proof. Let X, Y be the column vectors of our elements. Then $AX = Y$. Suppose Y is a basis. Then there exists a matrix C in R such that $CY = X$.

Then $CAX = X$, whence $CA = I$ and A is invertible. Conversely, assume that A is invertible. Then $X = A^{-1}Y$ and hence $x_1, \dots, x_n$ are in the module generated by $y_1, \dots, y_n$. Suppose that we have a relation

$$b_1 y_1 + \cdots + b_n y_n = 0$$

with $b_i \in R$. Let B be the row vector $(b_1, \dots, b_n)$. Then

$$BY = 0$$

and hence $BAX = 0$. But $\{x_1, \dots, x_n\}$ is a basis. Hence $BA = 0$, and hence $BAA^{-1} = B = 0$. This proves that the components of Y are linearly independent over R, and proves our proposition.

We return to our situation of modules over an arbitrary commutative ring R.

Let E, F be modules. We shall see how we can associate a matrix with a linear map whenever bases of E and F are given. We assume that E, F are free. We let $\mathcal{B} = \{\xi_1, \dots, \xi_n\}$ and $\mathcal{B}' = \{\xi'_1, \dots, \xi'_m\}$ be bases of E and F respectively. Let

$$f: E \to F$$

be a linear map. There exist unique elements $a_{ij} \in R$ such that

$$f(\xi_1) = a_{11}\xi'_1 + \cdots + a_{m1}\xi'_m,$$
$$\cdots$$
$$f(\xi_n) = a_{1n}\xi'_1 + \cdots + a_{mn}\xi'_m,$$

or in other words,

$$f(\xi_j) = \sum_{i=1}^{m} a_{ij}\xi'_i$$

(Observe that the sum is over the *first* index.) We define

$$M^{\mathcal{B}}_{\mathcal{B}'}(f) = (a_{ij}).$$

If $x = x_1\xi_1 + \cdots + x_n\xi_n$ is expressed in terms of the basis, let us denote the column vector X of components of x by $M_{\mathcal{B}}(x)$. We see that

$$M_{\mathcal{B}'}(f(x)) = M^{\mathcal{B}}_{\mathcal{B}'}(f)M_{\mathcal{B}}(x).$$

In other words, if X' is the column vector of $f(x)$, and M is the matrix associated with f then $X' = MX$. Thus the operation of the linear map is reflected by the matrix multiplication, and we have $f = L_M$.

Proposition 3.2. *Let E, F, D be modules, and let $\mathcal{B}$, $\mathcal{B}'$, $\mathcal{B}''$ be finite bases of E, F, D, respectively. Let*

$$E \xrightarrow{f} F \xrightarrow{g} D$$

be linear maps. Then

$$M^{\mathcal{B}}_{\mathcal{B}''}(g \circ f) = M^{\mathcal{B}'}_{\mathcal{B}''}(g) M^{\mathcal{B}}_{\mathcal{B}'}(f).$$

Proof. Let A and B be the matrices associated with the maps f, g respectively, with respect to our given bases. If X is the column vector associated with $x \in E$, the vector associated with $g(f(x))$ is $B(AX) = (BA)X$. Hence BA is the matrix associated with $g \circ f$. This proves what we wanted.

Corollary 3.3. *Let $E = F$. Then*

$$M^{\mathcal{B}}_{\mathcal{B}'}(\text{id}) M^{\mathcal{B}'}_{\mathcal{B}}(\text{id}) = M^{\mathcal{B}'}_{\mathcal{B}'}(\text{id}) = I.$$

Each matrix $M^{\mathcal{B}}_{\mathcal{B}'}(\text{id})$ is invertible (i.e. is a unit in the ring of matrices).

Proof. Obvious.

Corollary 3.4. *Let $N = M^{\mathcal{B}}_{\mathcal{B}'}(\text{id})$. Then*

$$M^{\mathcal{B}'}_{\mathcal{B}'}(f) = M^{\mathcal{B}}_{\mathcal{B}'}(\text{id}) M^{\mathcal{B}}_{\mathcal{B}}(f) M^{\mathcal{B}'}_{\mathcal{B}}(\text{id}) = N M^{\mathcal{B}}_{\mathcal{B}}(f) N^{-1}.$$

Proof. Obvious

Corollary 3.5. *Let E be a free module of dimension n over R. Let $\mathcal{B}$ be a basis of E over R. The map*

$$f \mapsto M^{\mathcal{B}}_{\mathcal{B}}(f)$$

is a ring-isomorphism of the ring of endomorphisms of E onto the ring of $n \times n$ matrices in R. In fact, the isomorphism is one of algebras over R.

We shall call the matrix $M^{\mathcal{B}}_{\mathcal{B}}(f)$ the **matrix associated with f with respect to the basis $\mathcal{B}$**.

Let E be a free module of dimension n over R. By $GL(E)$ or $\text{Aut}_R(E)$ one means the group of linear automorphisms of E. It is the group of units in $\text{End}_R(E)$. By $GL_n(R)$ one means the group of invertible $n \times n$ matrices in R. Once a basis is selected for E over R, we have a group-isomorphism

$$GL(E) \leftrightarrow GL_n(R)$$

with respect to this basis.

Let E be as above. If

$$f: E \to E$$

is a linear map, we select a basis $\mathcal{B}$ and let M be the matrix associated with f relative to $\mathcal{B}$. We define the **trace** of f to be the trace of M, thus

$$\operatorname{tr}(f) = \operatorname{tr}(M).$$

If M' is the matrix of f with respect to another basis, then there exists an invertible matrix N such that $M' = N^{-1}MN$, and hence the trace is independent of the choice of basis.

§4. DETERMINANTS

Let $E_1, \dots, E_n, F$ be modules. A map

$$f: E_1 \times \cdots \times E_n \to F$$

is said to be ***R*-multilinear** (or simply multilinear) if it is linear in each variable, i.e. if for every index i and elements $x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n, x_j \in E_j$, the map

$$x \mapsto f(x_1, \dots, x_{i-1}, x, x_{i+1}, \dots, x_n)$$

is a linear map of E_i into F.

A multilinear map defined on an n-fold product is also called n-multilinear. If $E_1 = \cdots = E_n = E$, we also say that f is a **multilinear map on *E***, instead of saying that it is multilinear on $E^{(n)}$.

Let f be an n-multilinear map. If we take two indices i, j and $i \neq j$ then fixing all the variables except the i-th and j-th variable, we can view f as a bilinear map on $E_i \times E_j$.

Assume that $E_1 = \cdots = E_n = E$. We say that the multilinear map f is **alternating** if $f(x_1, \dots, x_n) = 0$ whenever there exists an index i, $1 \leqq i \leqq n - 1$, such that $x_i = x_{i+1}$ (in other words, when two adjacent elements are equal).

Proposition 4.1. *Let f be an n-multilinear alternating map on E. Let $x_1, \dots, x_n \in E$. Then*

$$f(\dots, x_i, x_{i+1}, \dots) = -f(\dots, x_{i+1}, x_i, \dots).$$

In other words, when we interchange two adjacent arguments of f, the value of f changes by a sign. If $x_i = x_j$ for $i \neq j$ then $f(x_1, \dots, x_n) = 0$.

Proof. Restricting our attention to the factors in the i-th and j-th place, with $j = i + 1$, we may assume f is bilinear for the first statement. Then for all x, $y \in E$ we have

$$0 = f(x + y, x + y) = f(x, y) + f(y, x).$$

This proves what we want, namely $f(y, x) = -f(x, y)$. For the second assertion, we can interchange successively adjacent arguments of f until we obtain an n-tuple of elements of E having two equal adjacent arguments. This shows that when $x_i = x_j$, $i \neq j$, then $f(x_1, \dots, x_n) = 0$.

Corollary 4.2. *Let f be an n-multilinear alternating map on E. Let $x_1, \dots, x_n \in E$. Let $i \neq j$ and let $a \in R$. Then the value of f on $(x_1, \dots, x_n)$ does not change if we replace x_i by $x_i + ax_j$ and leave all other components fixed.*

Proof. Obvious.

A multilinear alternating map taking its value in R is called a multilinear alternating **form**.

On repeated occasions we shall evaluate multilinear alternating maps on linear combinations of elements of E. Let

$$w_1 = a_{11}v_1 + \cdots + a_{1n}v_n,$$
$$\cdots$$
$$w_n = a_{n1}v_1 + \cdots + a_{nn}v_n.$$

Let f be n-multilinear alternating on E. Then

$$f(w_1, \dots, w_n) = f(a_{11}v_1 + \cdots + a_{1n}v_n, \dots, a_{n1}v_1 + \cdots + a_{nn}v_n).$$

We expand this by multilinearity, and get a sum of terms of type

$$a_{1,\sigma(1)} \cdots a_{n,\sigma(n)} f(v_{\sigma(1)}, \dots, v_{\sigma(n)}),$$

where σ ranges over arbitrary maps of $\{1, \dots, n\}$ into itself. If σ is not a bijection (i.e. a permutation), then two arguments $v_{\sigma(i)}$ and $v_{\sigma(j)}$ are equal for $i \neq j$, and the term is equal to 0. Hence we may restrict our sum to permutations σ. Shuffling back the elements $(v_{\sigma(1)}, \dots, v_{\sigma(n)})$ to their standard ordering and using Proposition 4.1, we see that we have obtained the following expansion:

Lemma 4.3. *If $w_1, \dots, w_n$ are as above, then*

$$f(w_1, \dots, w_n) = \sum_{\sigma} \epsilon(\sigma) a_{1,\sigma(1)} \cdots a_{n,\sigma(n)} f(v_1, \dots, v_n)$$

where the sum is taken over all permutations σ of $\{1, \dots, n\}$ and $\epsilon(\sigma)$ is the sign of the permutation.

For determinants, I shall follow Artin's treatment in *Galois Theory.*

By an $n \times n$ **determinant** we shall mean a mapping

$$\det : \mathrm{Mat}_n(R) \to R$$

also written

$$D : \mathrm{Mat}_n(R) \to R$$

which, when viewed as a function of the column vectors $A^1, \ldots, A^n$ of a matrix A, is multilinear alternating, and such that $D(I) = 1$. In this chapter, we use mostly the letter D to denote determinants.

We shall prove later that determinants exist. For the moment, we derive properties.

Theorem 4.4. **(Cramer's Rule).** *Let $A^1, \ldots, A^n$ be column vectors of dimension n. Let $x_1, \ldots, x_n \in R$ be such that*

$$x_1 A^1 + \cdots + x_n A^n = B$$

for some column vector B. Then for each i we have

$$x_i D(A^1, \ldots, A^n) = D(A^1, \ldots, B, \ldots, A^n),$$

where B in this last line occurs in the i-th place.

Proof. Say $i = 1$. We expand

$$D(B, A^2, \ldots, A^n) = \sum_{j=1}^{n} x_j D(A^j, A^2, \ldots, A^n),$$

and use Proposition 4.1 to get what we want (all terms on the right are equal to 0 except the one having x_1 in it).

Corollary 4.5. *Assume that R is a field. Then $A^1, \ldots, A^n$ are linearly dependent if and only if $D(A^1, \ldots, A^n) = 0$.*

Proof. Assume we have a relation

$$x_1 A^1 + \cdots + x_n A^n = 0$$

with $x_i \in R$. Then $x_i D(A) = 0$ for all i. If some $x_i \neq 0$ then $D(A) = 0$. Conversely, assume that $A^1, \ldots, A^n$ are linearly independent. Then we can express the unit vectors $e^1, \ldots, e^n$ as linear combinations

$$e^1 = b_{11} A^1 + \cdots + b_{1n} A^n,$$

$$\cdots$$

$$e^n = b_{n1} A^1 + \cdots + b_{nn} A^n$$

with $b_{ij} \in R$. But

$$1 = D(e^1, \ldots, e^n).$$

Using a previous lemma, we know that this can be expanded into a sum of terms involving $D(A^1, \ldots, A^n)$, and hence $D(A)$ cannot be 0.

Proposition 4.6. *If determinants exist, they are unique. If $A^1, \ldots, A^n$ are the column vectors of dimension n, of the matrix $A = (a_{ij})$, then*

$$D(A^1, \ldots, A^n) = \sum_{\sigma} \epsilon(\sigma) a_{\sigma(1),1} \cdots a_{\sigma(n),n},$$

where the sum is taken over all permutations σ of $\{1, \ldots, n\}$, and $\epsilon(\sigma)$ is the sign of the permutation.

Proof. Let $e^1, \ldots, e^n$ be the unit vectors as usual. We can write

$$A^1 = a_{11}e^1 + \cdots + a_{n1}e^n,$$
$$\cdots$$
$$A^n = a_{1n}e^n + \cdots + a_{nn}e^n.$$

Therefore

$$D(A^1, \ldots, A^n) = \sum_{\sigma} \epsilon(\sigma) a_{\sigma(1),1} \cdots a_{\sigma(n),n}$$

by the lemma. This proves that the value of the determinant is uniquely determined and is given by the expected formula.

Corollary 4.7. *Let $\varphi : R \to R'$ be a ring-homomorphism into a commutative ring. If A is a square matrix in R, define φA to be the matrix obtained by applying φ to each component of A. Then*

$$\varphi(D(A)) = D(\varphi A).$$

Proof. Apply φ to the expression of Proposition 4.6.

Proposition 4.8. *If A is a square matrix in R then*

$$D(A) = D({}^tA).$$

Proof. In a product

$$a_{\sigma(1),1} \cdots a_{\sigma(n),n}$$

each integer k from 1 to n occurs precisely once among the integers $\sigma(1), \ldots, \sigma(n)$. Hence we can rewrite this product in the form

$$a_{1,\sigma^{-1}(1)} \cdots a_{n,\sigma^{-1}(n)}.$$

Since $\epsilon(\sigma) = \epsilon(\sigma^{-1})$, we can rewrite the sum in Proposition 4.6 in the form

$$\sum_{\sigma} \epsilon(\sigma^{-1}) a_{1,\sigma^{-1}(1)} \cdots a_{n,\sigma^{-1}(n)}.$$

In this sum, each term corresponds to a permutation σ. However, as σ ranges over all permutations, so does σ^{-1}. Hence our sum is equal to

$$\sum_{\sigma} \epsilon(\sigma) a_{1,\sigma(1)} \cdots a_{n,\sigma(n)},$$

which is none other than $D({}^tA)$, as was to be shown.

Corollary 4.9. *The determinant is multilinear and alternating with respect to the rows of a matrix.*

We shall now prove existence, and prove simultaneously one additional important property of determinants.

When $n = 1$, we define $D(a) = a$ for any $a \in R$.

Assume that we have proved the existence of determinants for all integers $< n$ $(n \geqq 2)$. Let A be an $n \times n$ matrix in R, $A = (a_{ij})$. We let A_{ij} be the $(n-1) \times (n-1)$ matrix obtained from A by deleting the i-th row and j-th column. Let i be a fixed integer, $1 \leqq i \leqq n$. We define inductively

$$D(A) = (-1)^{i+1} a_{i1} D(A_{i1}) + \cdots + (-1)^{i+n} a_{in} D(A_{in}).$$

(This is known as the **expansion of D according to the i-th row**.) We shall prove that D satisfies the definition of a determinant.

Consider D as a function of the k-th column, and consider any term

$$(-1)^{i+j} a_{ij} D(A_{ij}).$$

If $j \neq k$ then a_{ij} does not depend on the k-th column, and $D(A_{ij})$ depends linearly on the k-th column. If $j = k$, then a_{ij} depends linearly on the k-th column, and $D(A_{ij})$ does not depend on the k-th column. In any case our term depends linearly on the k-th column. Since $D(A)$ is a sum of such terms, it depends linearly on the k-th column, and thus D is multilinear.

Next, suppose that two adjacent columns of A are equal, say $A^k = A^{k+1}$. Let j be an index $\neq k$ and $\neq k+1$. Then the matrix A_{ij} has two adjacent equal columns, and hence its determinant is equal to 0. Thus the term corresponding to an index $j \neq k$ or $k+1$ gives a zero contribution to $D(A)$. The other two terms can be written

$$(-1)^{i+k} a_{ik} D(A_{ik}) + (-1)^{i+k+1} a_{i,k+1} D(A_{i,k+1}).$$

The two matrices A_{ik} and $A_{i,k+1}$ are equal because of our assumption that the k-th column of A is equal to the $(k+1)$-th column. Similarly, $a_{ik} = a_{i,k+1}$.

Hence these two terms cancel since they occur with opposite signs. This proves that our form is alternating, and gives:

Proposition 4.10. *Determinants exist and satisfy the rule of expansion according to rows and columns.*

(For columns, we use the fact that $D(A) = D({}^tA)$.)

Example. We mention explicity one of the most important determinants. Let $x_1, \dots, x_n$ be elements of a commutative ring. The **Vandermonde determinant** $V = V(x_1, \dots, x_n)$ of these elements is defined to be

$$V = \begin{vmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_n \\ \vdots & \vdots & & \vdots \\ x_1^{n-1} & x_2^{n-1} & \cdots & x_n^{n-1} \end{vmatrix},$$

whose value can be determined explicitly to be

$$V = \prod_{i<j} (x_j - x_i).$$

If the ring is entire and $x_i \neq x_j$ for $i \neq j$, it follows that $V \neq 0$. The proof for the stated value is done by multiplying the next to the last row by x_1 and subtracting from the last row. Then repeat this step going up the rows, thus making the elements of the first column equal to 0, except for 1 in the upper left-hand corner. One can then expand according to the first column, and use the homogeneity property and induction to conclude the proof of the evaluation of V.

Theorem 4.11. *Let E be a module over R, and let $v_1, \dots, v_n$ be elements of E. Let $A = (a_{ij})$ be a matrix in R, and let*

$$A\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}.$$

Let Δ be an n-multilinear alternating map on E. Then

$$\Delta(w_1, \dots, w_n) = D(A)\,\Delta(v_1, \dots, v_n).$$

Proof. We expand

$$\Delta(a_{11}v_1 + \cdots + a_{1n}v_n, \dots, a_{n1}v_1 + \cdots + a_{nn}v_n),$$

and find precisely what we want, taking into account $D(A) = D({}^tA)$.

Let E, F be modules, and let $L_a^n(E, F)$ denote the set of n-multilinear alternating maps of E into F. If $F = R$, we also write $L_a^n(E, R) = L_a^n(E)$. It is clear that $L_a^n(E, F)$ is a module over R, i.e. is closed under addition and multiplication by elements of R.

Corollary 4.12. *Let E be a free module over R, and let $\{v_1, \dots, v_n\}$ be a basis. Let F be any module, and let $w \in F$. There exists a unique n-multilinear alternating map*

$$\Delta_w : E \times \cdots \times E \to F$$

such that $\Delta_w(v_1, \dots, v_n) = w$.

Proof. Without loss of generality, we may assume that $E = R^{(n)}$, and then, if $A^1, \dots, A^n$ are column vectors, we define

$$\Delta_w(A^1, \dots, A^n) = D(A)w.$$

Then Δ_w obviously has the required properties.

Corollary 4.13. *If E is free over R, and has a basis consisting of n elements, then $L_a^n(E)$ is free over R, and has a basis consisting of 1 element.*

Proof. We let Δ_1 be the multilinear alternating map taking the value 1 on a basis $\{v_1, \dots, v_n\}$. Any element $\varphi \in L_a^n(E)$ can then be written in a unique way as $c\Delta_1$, with some $c \in R$, namely $c = \varphi(v_1, \dots, v_n)$. This proves what we wanted.

Any two bases of $L_a^n(E)$ in the preceding corollary differ by a unit in R. In other words, if Δ is a basis of $L_a^n(E)$, then $\Delta = c\Delta_1 = \Delta_c$ for some $c \in R$, and c must be a unit. Our Δ_1 depends of course on the choice of a basis for E. When we consider $R^{(n)}$, our determinant D is precisely Δ_1, relative to the standard basis consisting of the unit vectors $e^1, \dots, e^n$.

It is sometimes convenient terminology to say that any basis of $L_a^n(E)$ is a **determinant** on E. In that case, the corollary to Cramer's rule can be stated as follows.

Corollary 4.14. *Let R be a field. Let E be a vector space of dimension n. Let Δ be any determinant on E. Let $v_1, \dots, v_n \in E$. In order that $\{v_1, \dots, v_n\}$ be a basis of E it is necessary and sufficient that*

$$\Delta(v_1, \dots, v_n) \neq 0.$$

Proposition 4.15. *Let A, B be $n \times n$ matrices in R. Then*

$$D(AB) = D(A)D(B).$$

Proof. This is actually a corollary of Theorem 4.11. We take $v_1, \ldots, v_n$ to be the unit vectors $e^1, \ldots, e^n$, and consider

$$AB\begin{pmatrix} e^1 \\ \vdots \\ e^n \end{pmatrix} = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}.$$

We obtain

$$D(w_1, \ldots, w_n) = D(AB)D(e^1, \ldots, e^n).$$

On the other hand, by associativity, applying Theorem 4.11 twice,

$$D(w_1, \ldots, w_n) = D(A)D(B)D(e^1, \ldots, e^n).$$

Since $D(e^1, \ldots, e^n) = 1$, our proposition follows.

Let $A = (a_{ij})$ be an $n \times n$ matrix in R. We let

$$\tilde{A} = (b_{ij})$$

be the matrix such that

$$b_{ij} = (-1)^{i+j}D(A_{ji}).$$

(Note the reversal of indices!)

Proposition 4.16. *Let $d = D(A)$. Then $A\tilde{A} = \tilde{A}A = dI$. The determinant $D(A)$ is invertible in R if and only if A is invertible, and then*

$$A^{-1} = \frac{1}{d}\tilde{A}.$$

Proof. For any pair of indices i, k the ik-component of $A\tilde{A}$ is

$$a_{i1}b_{1k} + a_{i2}b_{2k} + \cdots + a_{in}b_{nk} = a_{i1}(-1)^{k+1}D(A_{k1}) + \cdots + a_{in}(-1)^{k+n}D(A_{kn}).$$

If $i = k$, then this sum is simply the expansion of the determinant according to the i-th row, and hence this sum is equal to d. If $i \neq k$, let $\bar{A}$ be the matrix obtained from A by replacing the k-th row by the i-th row, and leaving all other rows unchanged. If we delete the k-th row and the j-th column from $\bar{A}$, we obtain the same matrix as by deleting the k-th row and j-th column from A. Thus

$$\bar{A}_{kj} = A_{kj},$$

and hence our sum above can be written

$$a_{i1}(-1)^{k+1}D(\bar{A}_{k1}) + \cdots + a_{in}(-1)^{k+n}D(\bar{A}_{kn}).$$

This is the expansion of the determinant of $\bar{A}$ according to the i-th row. Hence $D(\bar{A}) = 0$, and our sum is 0. We have therefore proved that the ik-component of $A\tilde{A}$ is equal to d if $i = k$ (i.e. if it is a diagonal component), and is equal to 0 otherwise. This proves that $A\tilde{A} = dI$. On the other hand, we see at once from the definitions that ${}^t\tilde{A} = \widetilde{{}^tA}$. Then

$$ {}^t(\tilde{A}A) = {}^tA\,{}^t\tilde{A} = {}^tA\,\widetilde{{}^tA} = dI, $$

and consequently, $\tilde{A}A = dI$ also, since ${}^t(dI) = dI$. When d is a unit in R, then A is invertible, its inverse being $d^{-1}\tilde{A}$. Conversely, if A is invertible, and $AA^{-1} = I$, then $D(A)D(A^{-1}) = 1$, and hence $D(A)$ is invertible, as was to be shown.

Corollary 4.17. *Let F be any R-module, and let $w_1, \dots, w_n$ be elements of F. Let $A = (a_{ij})$ be an $n \times n$ matrix in R. Let*

$$ a_{11}w_1 + \cdots + a_{1n}w_n = v_1 $$
$$ \cdots $$
$$ a_{n1}w_1 + \cdots + a_{nn}w_n = v_n. $$

Then one can solve explicitly

$$ \begin{pmatrix} D(A)w_1 \\ \vdots \\ D(A)w_n \end{pmatrix} = D(A)\begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = \tilde{A}\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}. $$

In particular, if $v_i = 0$ for all i, then $D(A)w_i = 0$ for all i. If $v_i = 0$ for all i and F is generated by $w_1, \dots, w_n$, then $D(A)F = 0$.

Proof. This is immediate from the relation $\tilde{A}A = D(A)I$, using the remarks in §3 about applying matrices to column vectors whose components lie in the module.

Proposition 4.18. *Let E, F be free modules of dimension n over R. Let $f: E \to F$ be a linear map. Let $\mathcal{B}$, $\mathcal{B}'$ be bases of E, F respectively over R. Then f is an isomorphism if and only if the determinant of its associated matrix $M^{\mathcal{B}}_{\mathcal{B}'}(f)$ is a unit in R.*

Proof. Let $A = M^{\mathcal{B}}_{\mathcal{B}'}(f)$. By definition, f is an isomorphism if and only if there exists a linear map $g: F \to E$ such that $g \circ f = \text{id}$ and $f \circ g = \text{id}$. If f is an isomorphism, and $B = M^{\mathcal{B}'}_{\mathcal{B}}(g)$, then $AB = BA = I$. Taking the determinant of the product, we conclude that $D(A)$ is invertible in R. Conversely, if $D(A)$ is a unit, then we can define A^{-1} by Proposition 4.16. This A^{-1} is the associated matrix of a linear map $g: F \to E$ which is an inverse for f, as desired.

Finally, we shall define the determinant of an endomorphism.

Let E be a free module over R, and let $\mathcal{B}$ be a basis. Let $f: E \to E$ be an endomorphism of E. Let

$$M = M^{\mathcal{B}}_{\mathcal{B}}(f).$$

If $\mathcal{B}'$ is another basis of E, and $M' = M^{\mathcal{B}'}_{\mathcal{B}'}(f)$, then there exists an invertible matrix N such that

$$M' = NMN^{-1}.$$

Taking the determinant, we see that $D(M') = D(M)$. Hence the determinant does not depend on the choice of basis, and will be called the **determinant of the linear map** f. We shall give below a characterization of this determinant which does not depend on the choice of a basis.

Let E be any module. Then we can view $L^n_a(E)$ as a functor in the variable E (contravariant). In fact, we can view $L^n_a(E, F)$ as a functor of two variables, contravariant in the first, and covariant in the second. Indeed, suppose that

$$E' \xrightarrow{f} E$$

is a linear map. To each multilinear map $\varphi : E^{(n)} \to F$ we can associate the composite map $\varphi \circ f^{(n)}$,

$$E' \times \cdots \times E' \xrightarrow{f^{(n)}} E \times \cdots \times E \xrightarrow{\varphi} F$$

where $f^{(n)}$ is the product of f with itself n times. The map

$$L^n_a(f) : L^n_a(E, F) \to L^n_a(E', F)$$

given by

$$\varphi \mapsto \varphi \circ f^{(n)},$$

is obviously a linear map, which defines our functor. We shall sometimes write f^* instead of $L^n_a(f)$.

In particular, consider the case when $E = E'$ and $F = R$. We get an induced map

$$f^* : L^n_a(E) \to L^n_a(E).$$

Proposition 4.19. *Let E be a free module over R, of dimension n. Let $\{\Delta\}$ be a basis of $L^n_a(E)$. Let $f: E \to E$ be an endomorphism of E. Then*

$$f^*\Delta = D(f)\Delta.$$

Proof. This is an immediate consequence of Theorem 4.11. Namely, we let $\{v_1, \dots, v_n\}$ be a basis of E, and then take A (or ${}^tA$) to be a matrix of f relative to this basis. By definition,

$$f^*\Delta(v_1, \dots, v_n) = \Delta(f(v_1), \dots, f(v_n)),$$

and by Theorem 4.11, this is equal to

$$D(A)\,\Delta(v_1, \ldots, v_n).$$

By Corollary 4.12, we conclude that $f^*\Delta = D(A)\Delta$ since both of these forms take on the same value on $(v_1, \ldots, v_n)$.

The above considerations have dealt with the determinant as a function on all endomorphisms of a free module. One can also view it multiplicatively, as a homomorphism.

$$\det : GL_n(R) \to R^*$$

from the group of invertible $n \times n$ matrices over R into the group of units of R. The kernel of this homomorphism, consisting of those matrices with determinant 1, is called the **special linear group**, and is denoted by $SL_n(R)$.

We now give an application of determinants to the situation of a free module and a submodule considered in Chapter III, Theorem 7.8.

Proposition 4.20. *Let R be a principal entire ring. Let F be a free module over R and let M be a finitely generated submodule. Let $\{e_1, \ldots, e_m, \ldots\}$ be a basis of F such that there exist non-zero elements $a_1, \ldots, a_m \in R$ such that:*

(i) *The elements $a_1e_1, \ldots, a_me_m$ form a basis of M over R.*
(ii) *We have $a_i \mid a_{i+1}$ for $i = 1, \ldots, m - 1$.*

Let L_a^s be the set of all s-multilinear alternating forms on F. Let J_s be the ideal generated by all elements $f(y_1, \ldots, y_s)$, with $f \in L_a^s$ and $y_1, \ldots, y_s \in M$. Then

$$J_s = (a_1 \cdots a_s).$$

Proof. We first show that $J_s \subset (a_1 \cdots a_s)$. Indeed, an element $y \in M$ can be written in the form

$$y = c_1a_1e_1 + \cdots + c_ra_re_r.$$

Hence if $y_1, \ldots, y_s \in M$, and f is multilinear alternating on F, then $f(y_1, \ldots, y_s)$ is equal to a sum in terms of type

$$c_{i_1} \cdots c_{i_s} a_{i_1} \cdots a_{i_s} f(e_{i_1}, \ldots, e_{i_s}).$$

This is non-zero only when $e_{i_1}, \ldots, e_{i_s}$ are distinct, in which case the product $a_1 \cdots a_s$ divides this term, and hence J_s is contained in the stated ideal.

Conversely, we show that there exists an s-multilinear alternating form which gives precisely this product. We deduce this from determinants. We can write F as a direct sum

$$F = (e_1, \ldots, e_r) \oplus F_r$$

with some submodule F_r. Let f_i $(i = 1, \dots, r)$ be the linear map $F \to R$ such that $f_i(e_j) = \delta_{ij}$, and such that f_i has value 0 on F_r. For $v_1, \dots, v_s \in F$ we define

$$f(v_1, \dots, v_s) = \det(f_i(v_j)).$$

Then f is multilinear alternating and takes on the value

$$f(e_2, \dots, e_s) = 1,$$

as well as the value

$$f(a_1 e_1, \dots, a_s e_s) = a_1 \cdots a_s.$$

This proves the proposition.

The uniqueness of Chapter III, Theorem 7.8 is now obvious, since first (a_1) is unique, then $(a_1 a_2)$ is unique and the quotient (a_2) is unique, and so forth by induction.

Remark. Compare the above theorem with Theorem 2.9 of Chapter XIX, in the theory of Fitting ideals, which gives a fancier context for the result.

---

§5. DUALITY

Let R be a commutative ring, and let E, F be modules over R. An ***R*-bilinear form** on $E \times F$ is a map

$$f: E \times F \to R$$

having the following properties: For each $x \in E$, the map

$$y \mapsto f(x, y)$$

is R-linear, and for each $y \in F$, the map

$$x \mapsto f(x, y)$$

is R-linear. We shall omit the prefix R- in the rest of this section, and write $\langle x, y \rangle_f$ or $\langle x, y \rangle$ instead of $f(x, y)$. If $x \in F$, we write $x \perp y$ if $\langle x, y \rangle = 0$. Similarly, if S is a subset of F, we define $x \perp S$ if $x \perp y$ for all $y \in S$. We then say that x is **perpendicular** to S. We let $S^{\perp}$ consist of all elements of E which are perpendicular to S. It is obviously a submodule of E. We define perpendicularity on the other side in the same way. We define the **kernel** of f on the left to be $F^{\perp}$ and the kernel on the right to be $E^{\perp}$. We say that f is **non-degenerate** on the left if its kernel on the left is 0. We say that f is **non-degenerate** on the right if its kernel on the right is 0. If E_0 is the kernel of f on the left, then we

get an induced bilinear map

$$E/E_0 \times F \to R$$

which is non-degenerate on the left, as one verifies trivially from the definitions. Similarly, if F_0 is the kernel of f on the right, we get an induced bilinear map

$$E/E_0 \times F/F_0 \to R$$

which is non-degenerate on either side. This map arises from the fact that the value $\langle x, y \rangle$ depends only on the coset of x modulo E_0 and the coset of y modulo F_0.

We shall denote by $L^2(E, F; R)$ the set of all bilinear maps of $E \times F$ into R. It is clear that this set is a module (i.e. an R-module), addition of maps being the usual one, and also multiplication of maps by elements of R.

The form f gives rise to a homomorphism

$$\varphi_f : E \to \mathrm{Hom}_R(F, R)$$

such that

$$\varphi_f(x)(y) = f(x, y) = \langle x, y \rangle,$$

for all $x \in E$ and $y \in F$. We shall call $\mathrm{Hom}_R(F, R)$ the **dual module** of F, and denote it by $F^\vee$. We have an *isomorphism*

$$\boxed{L^2(E, F; R) \leftrightarrow \mathrm{Hom}_R(E, \mathrm{Hom}_R(F, R))}$$

given by $f \mapsto \varphi_f$, its inverse being defined in the obvious way: If

$$\varphi : E \to \mathrm{Hom}_R(F, R)$$

is a homomorphism, we let f be such that

$$f(x, y) = \varphi(x)(y).$$

We shall say that f is **non-singular on the left** if φ_f is an isomorphism, in other words if our form can be used to identify E with the dual module of F. We define **non-singular on the right** in a similar way, and say that f is **non-singular** if it is non-singular on the left and on the right.

Warning: Non-degeneracy does not necessarily imply non-singularity.

We shall now obtain an *isomorphism*

$$\boxed{\mathrm{End}_R(E) \mapsto L^2(E, F; R)}$$

depending on a fixed non-singular bilinear map $f : E \times F \to R$.

Let $A \in \text{End}_R(E)$ be a linear map of E into itself. Then the map

$$(x, y) \mapsto \langle Ax, y \rangle = \langle Ax, y \rangle_f$$

is bilinear, and in this way, we associate linearly with each $A \in \text{End}_R(E)$ a bilinear map in $L^2(E, F; R)$.

Conversely, let $h : E \times F \to R$ be bilinear. Given $x \in E$, the map $h_x : F \to R$ such that $h_x(y) = h(x, y)$ is linear, and is in the dual space $F^\vee$. By assumption, there exists a unique element $x' \in E$ such that for all $y \in F$ we have

$$h(x, y) = \langle x', y \rangle.$$

It is clear that the association $x \mapsto x'$ is a linear map of E into itself. Thus with each bilinear map $E \times F \to R$ we have associated a linear map $E \to E$.

It is immediate that the mappings described in the last two paragraphs are inverse isomorphisms between $\text{End}_R(E)$ and $L^2(E, F; R)$. We emphasize of course that they depend on our form f.

Of course, we could also have worked on the right, and thus we have a similar *isomorphism*

$$\boxed{L^2(E, F; R) \leftrightarrow \text{End}_R(F)}$$

depending also on our fixed non-singular form f.

As an application, let $A : E \to E$ be linear, and let $(x, y) \mapsto \langle Ax, y \rangle$ be its associated bilinear map. There exists a unique linear map

$${}^tA : F \to F$$

such that

$$\langle Ax, y \rangle = \langle x, {}^tAy \rangle$$

for all $x \in E$ and $y \in F$. We call ${}^tA$ the **transpose of A with respect to f**.

It is immediately clear that if, A, B are linear maps of E into itself, then for $c \in R$,

$${}^t(cA) = c\,{}^tA, \qquad {}^t(A + B) = {}^tA + {}^tB, \quad \text{and} \quad {}^t(AB) = {}^tB\,{}^tA.$$

More generally, let E, F be modules with non-singular bilinear forms denoted by $\langle\ ,\ \rangle_E$ and $\langle\ ,\ \rangle_F$ respectively. Let $A: E \to F$ be a linear map. Then by the non-singularity of $\langle\ ,\ \rangle_E$ there exists a unique linear map ${}^tA: F \to E$ such that

$$\langle Ax, y \rangle_F = \langle x, {}^tAy \rangle_E \text{ for all } x \in E \text{ and } y \in F.$$

We also call ${}^tA$ the **transpose** with respect to these forms.

Examples. For a nice classical example of a transpose, see Exercise 33. For the systematic study when a linear map is equal to its transpose, see the

spectral theorems of Chapter XV. Next I give another example of a transpose from analysis as follows. Let E be the (infinite dimensional) vector space of C^∞ functions on $\mathbf{R}$, having compact support, i.e. equal to 0 outside some finite interval. We define the scalar product

$$\langle f, g\rangle = \int_{-\infty}^{\infty} f(x)g(x)dx.$$

Let $D: E \to E$ be the derivative. Then one has the formula

$$\langle Df, g\rangle = -\langle f, Dg\rangle.$$

Thus one says that ${}^tD = -D$, even though the scalar product is not "non-singular", but much of the formalism of non-singular forms goes over. Also in analysis, one puts various norms on the spaces and one extends the bilinear form by continuity to the completions, thus leaving the domain of algebra to enter the domain of estimates (analysis). Then the spectral theorems become more complicated in such analytic contexts.

Let us assume that $E = F$. Let $f: E \times E \to R$ be bilinear. By an **automorphism of the pair (E,f)**, or simply of f, we shall mean a linear automorphism $A: E \to E$ such that

$$\langle Ax, Ay\rangle = \langle x, y\rangle$$

for all $x, y \in E$. The group of automorphisms of f is denoted by $\mathrm{Aut}(f)$.

Proposition 5.1. *Let $f: E \times E \to R$ be a non-singular bilinear form. Let $A: E \to E$ be a linear map. Then A is an automorphism of f if and only if ${}^tAA = \mathrm{id}$, and A is invertible.*

Proof. From the equality

$$\langle x, y\rangle = \langle Ax, Ay\rangle = \langle x, {}^tAAy\rangle$$

holding for all $x, y \in E$, we conclude that ${}^tAA = \mathrm{id}$ if A is an automorphism of f. The converse is equally clear.

Note. If E is free and finite dimensional, then the condition ${}^tAA = \mathrm{id}$ implies that A is invertible.

Let $f: E \times E \to R$ be a bilinear form. We say that f is **symmetric** if $f(x, y) = f(y, x)$ for all $x, y \in E$. The set of symmetric bilinear forms on E will be denoted by $L_s^2(E)$. Let us take a fixed symmetric non-singular bilinear form f on E, denoted by $(x, y) \mapsto \langle x, y\rangle$. An endomorphism $A: E \to E$ will be said to be **symmetric with respect** to f if ${}^tA = A$. It is clear that the set of symmetric endomorphisms of E is a module, which we shall denote by $\mathrm{Sym}(E)$.

Depending on our fixed symmetric non-singular f, we have an isomorphism

$$\boxed{L_s^2(E) \leftrightarrow \mathrm{Sym}(E)}$$

which we describe as follows. If g is symmetric bilinear on E, then there exists a unique linear map A such that

$$g(x, y) = \langle Ax, y\rangle$$

for all $x, y \in E$. Using the fact that both f, g are symmetric, we obtain

$$\langle Ax, y\rangle = \langle Ay, x\rangle = \langle y, {}^tAx\rangle = \langle {}^tAx, y\rangle.$$

Hence $A = {}^tA$. The association $g \mapsto A$ gives us a homomorphism from $L_s^2(E)$ into $\mathrm{Sym}(E)$. Conversely, given a symmetric endomorphism A of E, we can define a symmetric form by the rule $(x, y) \mapsto \langle Ax, y\rangle$, and the association of this form to A clearly gives a homomorphism of $\mathrm{Sym}(E)$ into $L_s^2(E)$ which is inverse to the preceding homomorphism. Hence $\mathrm{Sym}(E)$ and $L_s^2(E)$ are isomorphic.

We recall that a bilinear form $g : E \times E \to R$ is said to be **alternating** if $g(x, x) = 0$ for all $x \in E$, and consequently $g(x, y) = -g(y, x)$ for all $x, y \in E$. The set of bilinear alternating forms on E is a module, denoted by $L_a^2(E)$.

Let f be a fixed *symmetric* non-singular bilinear form on E. An endomorphism $A : E \to E$ will be said to be **skew-symmetric** or **alternating** with respect to f if ${}^tA = -A$, and also $\langle Ax, x\rangle = 0$ for all $x \in E$. If for all $a \in R$, $2a = 0$ implies $a = 0$, then this second condition $\langle Ax, x\rangle = 0$ is redundant, because $\langle Ax, x\rangle = -\langle Ax, x\rangle$ implies $\langle Ax, x\rangle = 0$. It is clear that the set of alternating endomorphisms of E is a module, denoted by $\mathrm{Alt}(E)$. *Depending on our fixed symmetric non-singular form f, we have an isomorphism*

$$\boxed{L_a^2(E) \leftrightarrow \mathrm{Alt}(E)}$$

described as usual. If g is an alternating bilinear form on E, its corresponding linear map A is the one such that

$$g(x, y) = \langle Ax, y\rangle$$

for all $x, y \in E$. One verifies trivially in a manner similar to the one used in the symmetric case that the correspondence $g \leftrightarrow A$ gives us our desired isomorphism.

Examples. Let k be a field and let E be a finite-dimensional vector space over k. Let $f : E \times E \to E$ be a bilinear map, denoted by $(x, y) \mapsto xy$. To each

$x \in E$, we associate the linear map $\lambda_x : E \mapsto E$ such that

$$\lambda_x(y) = xy.$$

Then the map obtained by taking the trace, namely

$$(x, y) \mapsto \operatorname{tr}(\lambda_{xy})$$

is a bilinear form on E. If $xy = yx$, then this bilinear form is symmetric.

Next, let E be the space of continuous functions on the interval $[0, 1]$. Let $K(s, t)$ be a continuous function of two real variables defined on the square $0 \leqq s \leqq 1$ and $0 \leqq t \leqq 1$. For $\varphi, \psi \in E$ we define

$$\langle \varphi, \psi \rangle = \iint \varphi(s) K(s, t) \psi(t) \, ds \, dt,$$

the double integral being taken on the square. Then we obtain a bilinear form on E. If $K(s, t) = K(t, s)$, then the bilinear form is symmetric. When we discuss matrices and bilinear forms in the next section, the reader will note the similarity between the preceding formula and the bilinear form defined by a matrix.

Thirdly, let U be an open subset of a real Banach space E (or a finite-dimensional Euclidean space, if the reader insists), and let $f: U \to \mathbf{R}$ be a map which is twice continuously differentiable. For each $x \in U$, the derivative $Df(x): E \to \mathbf{R}$ is a continuous linear map, and the second derivative $D^2f(x)$ can be viewed as a continuous symmetric bilinear map of $E \times E$ into $\mathbf{R}$.

§6. MATRICES AND BILINEAR FORMS

We shall investigate the relation between the concepts introduced above and matrices. Let $f: E \times F \to R$ be bilinear. Assume that E, F are free over R. Let $\mathfrak{B} = \{v_1, \ldots, v_m\}$ be a basis for E over R, and let $\mathfrak{B}' = \{w_1, \ldots, w_n\}$ be a basis for F over R. Let $g_{ij} = \langle v_i, w_j \rangle$. If

$$x = x_1 v_1 + \cdots + x_m v_m$$

and

$$y = y_1 w_1 + \cdots + y_n w_n$$

are elements of E and F respectively, with coordinates $x_i, y_j \in R$, then

$$\langle x, y \rangle = \sum_{i=1}^{m} \sum_{j=1}^{n} g_{ij} x_i y_j.$$

Let X, Y be the column vectors of coordinates for x, y respectively, with respect to our bases. Then

$$\langle x, y \rangle = {}^tXGY$$

where G is the matrix (g_{ij}). We could write $G = M^{\mathcal{B}}_{\mathcal{B}'}(f)$. We call G the **matrix associated with the form f relative to the bases $\mathcal{B}$, $\mathcal{B}'$**.

Conversely, given a matrix G (of size $m \times n$), we get a bilinear form from the map

$$(X, Y) \mapsto {}^tXGY.$$

In this way, we get a correspondence from bilinear forms to matrices and back, and it is clear that this correspondence induces an *isomorphism* (of R-modules)

$$\boxed{L^2(E, F; R) \leftrightarrow \mathrm{Mat}_{m \times n}(R)}$$

given by

$$f \mapsto M^{\mathcal{B}}_{\mathcal{B}'}(f).$$

The two maps between these two modules which we described above are clearly inverse to each other.

If we have bases $\mathcal{B} = \{v_1, \ldots, v_n\}$ and $\mathcal{B}' = \{w_1, \ldots, w_n\}$ such that $\langle v_i, w_j \rangle = \delta_{ij}$, then we say that these bases are **dual** to each other. In that case, if X is the coordinate vector of an element of E, and Y the coordinate vector of an element of F, then the bilinear map on X, Y has the value

$$X \cdot Y = x_1 y_1 + \cdots + x_n y_n$$

given by the usual dot product.

It is easy to derive in general how the matrix G changes when we change bases in E and F. However, we shall write down the explicit formula only when $E = F$ and $\mathcal{B} = \mathcal{B}'$. Thus we have a bilinear form $f: E \times E \to R$. Let $\mathcal{C}$ be another basis of E and write $X_{\mathcal{B}}$ and $X_{\mathcal{C}}$ for the column vectors belonging to an element x of E, relative to the two bases. *Let C be the invertible matrix* $M^{\mathcal{C}}_{\mathcal{B}}(\mathrm{id})$, so that

$$X_{\mathcal{B}} = CX_{\mathcal{C}}.$$

Then our form is given by

$$\langle x, y \rangle = {}^tX_{\mathcal{C}}\,{}^tCGCY_{\mathcal{C}}.$$

We see that

$$M^{\mathcal{C}}_{\mathcal{C}}(f) = {}^tCM^{\mathcal{B}}_{\mathcal{B}}(f)C. \tag{1}$$

In other words, the matrix of the bilinear form changes by the *transpose*.

If F is free over R, with a basis $\{\eta_1, \ldots, \eta_n\}$, then $\mathrm{Hom}_R(F, R)$ *is also free, and we have a dual basis $\{\eta'_1, \ldots, \eta'_n\}$ such that*

$$\eta'_i(\eta_j) = \delta_{ij}.$$

This has already been mentioned in Chapter III, Theorem 6.1.

Proposition 6.1. *Let E, F be free modules of dimension n over R and let $f: E \times F \to R$ be a bilinear form. Then the following conditions are equivalent:*

f is non-singular on the left.
f is non-singular on the right.
f is non-singular.
The determinant of the matrix of f relative to any bases is invertible in R.

Proof. Assume that f is non-singular on the left. Fix bases of E and F relative to which we write elements of these modules as column vectors, and giving rise to the matrix G for f. Then our form is given by

$$(X, Y) \mapsto {}^tXGY$$

where X, Y are column vectors with coefficients in R. By assumption the map

$$X \mapsto {}^tXG$$

gives an isomorphism between the module of column vectors, and the module of row vectors of length n over R. Hence G is invertible, and hence its determinant is a unit in R. The converse is equally clear, and if $\det(G)$ is a unit, we see that the map

$$Y \to GY$$

must also be an isomorphism between the module of column vectors and itself. This proves our assertion.

We shall now investigate how the transpose behaves in terms of matrices. Let E, F be free over R, of dimension n.

Let $f: E \times F \to R$ be a non-singular bilinear form, and assume given a basis $\mathfrak{B}$ of E and $\mathfrak{B}'$ of F. Let G be the matrix of f relative to these bases. Let $A: E \to E$ be a linear map. If $x \in E$, $y \in F$, let X, Y be their column vectors relative to $\mathfrak{B}$, $\mathfrak{B}'$. Let M be the matrix of A relative to $\mathfrak{B}$. Then for $x \in E$ and $y \in F$ we have

$$\langle Ax, y\rangle = {}^t(MX)GY = {}^tX{}^tMGY.$$

Let N be the matrix of ${}^tA$ relative to the basis $\mathfrak{B}'$. Then NY is the column vector of ${}^tAy$ relative to $\mathfrak{B}'$. Hence

$$\langle x, {}^tAy\rangle = {}^tXGNY.$$

From this we conclude that ${}^tMG = GN$, and since G is invertible, we can solve for N in terms of M. We get:

Proposition 6.2. *Let E, F be free over R, of dimension n. Let $f: E \times F \to R$ be a non-singular bilinear form. Let $\mathcal{B}$, $\mathcal{B}'$ be bases of E and F respectively over R, and let G be the matrix of f relative to these bases. Let $A: E \to E$ be a linear map, and let M be its matrix relative to $\mathcal{B}$. Then the matrix of ${}^tA$ relative to $\mathcal{B}'$ is*

$$(G^{-1})^t MG.$$

Corollary 6.3. *If G is the unit matrix, then the matrix of the transpose is equal to the transpose of the matrix.*

In terms of matrices and bases, we obtain the following characterization for a matrix to induce an automorphism of the form.

Corollary 6.4. *Let the notation be as in Proposition 6.2, and let $E = F$, $\mathcal{B} = \mathcal{B}'$. An $n \times n$ matrix M is the matrix of an automorphism of the form f (relative to our basis) if and only if*

$${}^tMGM = G.$$

If this condition is satisfied, then in particular, M is invertible.

Proof. We use the definitions, together with the formula given in Proposition 6.2. We note that M is invertible, for instance because its determinant is a unit in R.

A matrix M is said to be **symmetric** (resp. **alternating**) if ${}^tM = M$ (resp. ${}^tM = -M$ and the diagonal elements of M are 0).

Let $f: E \times E \to R$ be a bilinear form. We say that f is **symmetric** if $f(x, y) = f(y, x)$ for all $x, y \in E$. We say that f is **alternating** if $f(x, x) = 0$ for all $x \in E$.

Proposition 6.5. *Let E be a free module of dimension n over R, and let $\mathcal{B}$ be a fixed basis. The map*

$$f \mapsto M^{\mathcal{B}}_{\mathcal{B}}(f)$$

induces an isomorphism between the module of symmetric bilinear forms on $E \times E$ (resp. the module of alternating forms on $E \times E$) and the module of symmetric $n \times n$ matrices over R (resp. the module of alternating $n \times n$ matrices over R).

Proof. Consider first the symmetric case. Assume that f is symmetric. In terms of coordinates, let $G = M_{\mathfrak{B}}^{\mathfrak{B}}(f)$. Our form is given by ${}^tXGY$ which must be equal to ${}^tYGX$ by symmetry. However, ${}^tXGY$ may be viewed as a 1×1 matrix, and is equal to its transpose, namely ${}^tY{}^tGX$. Thus

$$ {}^tYGX = {}^tY{}^tGX $$

for all vectors X, Y. It follows that $G = {}^tG$. Conversely, it is clear that any symmetric matrix defines a symmetric form.

As for the alternating case, replacing x by $x + y$ in the relation $\langle x, x \rangle = 0$ we obtain

$$ \langle x, y \rangle = \langle y, x \rangle = 0. $$

In terms of the coordinate vectors X, Y and the matrix G, this yields

$$ {}^tXGY + {}^tYGX = 0. $$

Taking the transpose of, say, the second of the 1×1 matrices entering in this relation, yields (for all X, Y):

$$ {}^tXGY + {}^tX{}^tGY = 0. $$

Hence $G + {}^tG = 0$. Furthermore, letting X be any one of the unit vectors

$$ {}^t(0, \ldots, 0, 1, 0, \ldots, 0) $$

and using the relation ${}^tXGX = 0$, we see that the diagonal elements of G must be equal to 0. Conversely, if G is an $n \times n$ matrix such that ${}^tG + G = 0$, and such that $g_{ii} = 0$ for $i = 1, \ldots, n$ then one verifies immediately that the map

$$ (X, Y) \mapsto {}^tXGY $$

defines an alternating form. This proves our proposition.

Of course, if as is usually the case, 2 is invertible in R, then our condition ${}^tM = -M$ implies that the diagonal elements of M must be 0. Thus in that case, showing that $G + {}^tG = 0$ implies that G is alternating.

§7. SESQUILINEAR DUALITY

There exist forms which are not quite bilinear, and for which the results described above hold almost without change, but which must be handled separately for the sake of clarity in the notation involved.

Let R have an automorphism of period 2. We write this automorphism as $a \mapsto \bar{a}$ (and think of complex conjugation).

Following Bourbaki, we say that a map

$$f : E \times F \to R$$

is a **sesquilinear form** if it is **Z**-bilinear, and if for $x \in E$, $y \in F$, and $a \in R$ we have

$$f(ax, y) = af(x, y)$$

and

$$f(x, ay) = \bar{a}f(x, y).$$

(**Sesquilinear** means $1\frac{1}{2}$ times linear, so the terminology is rather good.)

Let E, E' be modules. A map $\varphi : E \to E'$ is said to be **anti-linear** (or **semi-linear**) if it is **Z**-linear, and $\varphi(ax) = \bar{a}\varphi(x)$ for all $x \in E$. Thus we may say that a sesquilinear form is linear in its first variable, and anti-linear in its second variable. We let $\overline{\text{Hom}}_R(E, E')$ denote the module of anti-linear maps of E into E'.

We shall now go systematically through the same remarks that we made previously for bilinear forms.

We define perpendicularity as before, and also the kernel on the right and on the left for any sesquilinear form f. These kernels are submodules, say E_0 and F_0, and we get an induced sesquilinear form

$$E/E_0 \times F/F_0 \to R,$$

which is non-degenerate on either side.

Let F be an R-module. We define its **anti-module** $\bar{F}$ to be the module whose additive group is the same as F, and such that the operation $R \times \bar{F} \to \bar{F}$ is given by

$$(a, y) \mapsto \bar{a}y.$$

Then $\bar{F}$ is a module. We have a natural isomorphism

$$\text{Hom}_R(\bar{F}, R) \leftrightarrow \overline{\text{Hom}}_R(F, R),$$

as R-modules.

The sesquilinear form $f : E \times F \to R$ induces a linear map

$$\varphi_f : E \to \text{Hom}_R(\bar{F}, R).$$

We say that f is **non-singular on the left** if φ_f is an isomorphism. Similarly, we have a corresponding linear map

$$\varphi'_f : \bar{F} \to \text{Hom}_R(E, R)$$

from $\bar{F}$ into the dual space of E, and we say that f is **non-singular on the right** if φ'_f is an isomorphism. We say that f is **non-singular** if it is non-singular on the left and on the right.

We observe that our sesquilinear form f can be viewed as a **bilinear** form

$$f : E \times \bar{F} \to R,$$

and that our notions of non-singularity are then compatible with those defined previously for bilinear forms.

If we have a fixed non-singular sesquilinear form on $E \times F$, then depending on this form, we obtain an isomorphism between the module of sesquilinear forms on $E \times F$ and the module of endomorphisms of E. We also obtain an anti-isomorphism between these modules and the module of endomorphisms of F. In particular, we can define the analogue of the transpose, which in the present case we shall call the adjoint. Thus, let $f : E \times F \to R$ be a non-singular sesquilinear form. Let $A : E \to E$ be a linear map. There exists a unique linear map

$$A^* : F \to F$$

such that

$$\langle Ax, y \rangle = \langle x, A^*y \rangle$$

for all $x \in E$ and $y \in F$. Note that A^* is linear, not anti-linear. We call A^* the **adjoint** of A with respect to our form f. We have the rules

$$(cA)^* = \bar{c}A^*, \qquad (A + B)^* = A^* + B^*, \qquad (AB)^* = B^*A^*$$

for all linear maps A, B of E into itself, and $c \in R$.

Let us assume that $E = F$. Let $f : E \times E \to R$ be sesquilinear. By an **automorphism** of f we shall mean a linear automorphism $A : E \to E$ such that

$$\langle Ax, Ay \rangle = \langle x, y \rangle$$

just as we did for bilinear forms.

Proposition 7.1. *Let $f : E \times E \to R$ be a non-singular sesquilinear form. Let $A : E \to E$ be a linear map. Then A is an automorphism of f if and only if $A^*A = \mathrm{id}$, and A is invertible.*

The proof, and also the proofs of subsequent propositions, which are completely similar to those of the bilinear case, will be omitted.

A sesquilinear form $g : E \times E \to R$ is said to be **hermitian** if

$$g(x, y) = \overline{g(y, x)}$$

for all $x, y \in E$. The set of hermitian forms on E will be denoted by $L^2_h(E)$. Let R_0 be the subring of R consisting of all elements fixed under our automorphism

$a \to \bar{a}$ (i.e. consisting of all elements $a \in R$ such that $a = \bar{a}$). Then $L_h^2(E)$ is an R_0-module.

Let us take a fixed hermitian non-singular form f on E, denoted by $(x, y) \mapsto \langle x, y \rangle$. An endomorphism $A : E \to E$ will be said to be **hermitian** with respect to f if $A^* = A$. It is clear that the set of hermitian endomorphisms is an R_0-module, which we shall denote by Herm(E). *Depending on our fixed hermitian non-singular form f, we have an R_0-isomorphism*

$$\boxed{L_h^2(E) \leftrightarrow \text{Herm}(E)}$$

described in the usual way. A hermitian form g corresponds to a hermitian map A if and only if

$$g(x, y) = \langle Ax, y \rangle$$

for all $x, y \in E$.

We can now describe the relation between our concepts and matrices, just as we did with bilinear forms.

We start with a sesquilinear form $f : E \times F \to R$.

If E, F are free, and we have selected bases as before, then we can again associate a matrix G with the form, and in terms of coordinate vectors X, Y our sesquilinear form is given by

$$(X, Y) \mapsto {}^tXG\bar{Y},$$

where $\bar{Y}$ is obtained from Y by applying the automorphism to each component of Y.

If $E = F$ and we use the same basis on the right and on the left, then with the same notation as that used in formula (1), if f is sesquilinear, the formula now reads

$$M_{\mathscr{C}}^{\mathscr{C}}(f) = {}^tCM_{\mathscr{B}}^{\mathscr{B}}(f)\bar{C}. \tag{1S}$$

The automorphism appears.

Proposition 7.2. *Let E, F be free modules of dimension n over R, and let $f : E \times F \to R$ be a sesquilinear form. Then the following conditions are equivalent.*

f is non-singular on the left.
f is non-singular on the right.
f is non-singular.

The determinant of the matrix of f relative to any bases is invertible in R.

Proposition 7.3. *Let E, F be free over R, of dimension n. Let $f: E \times F \to R$ be a non-singular sesquilinear form. Let $\mathcal{B}$, $\mathcal{B}'$ be bases of E and F respectively over R, and let G be the matrix of f relative to these bases. Let $A: E \to E$ be a linear map, and let M be its matrix relative to $\mathcal{B}$. Then the matrix of A^* relative to $\mathcal{B}'$ is*

$$(\bar{G}^{-1})^t \bar{M} \bar{G}.$$

Corollary 7.4. *If G is the unit matrix, then the matrix of A^* is equal to ${}^t\bar{M}$.*

Corollary 7.5. *Let the notation be as in the proposition, and let $\mathcal{B} = \mathcal{B}'$ be a basis of E. An $n \times n$ matrix M is the matrix of an automorphism of f (relative to our basis) if and only if*

$${}^tMG\bar{M} = G.$$

A matrix M is said to be **hermitian** if ${}^tM = \bar{M}$.

Let R_0 be as before the subring of R consisting of all elements fixed under our automorphism $a \mapsto \bar{a}$ (i.e. consisting of all elements $a \in R$ such that $a = \bar{a}$).

Proposition 7.6. *Let E be a free module of dimension n over R, and let $\mathcal{B}$ be a basis. The map*

$$f \mapsto M_{\mathcal{B}}^{\mathcal{B}}(f)$$

induces an R_0-isomorphism between the R_0-module of hermitian forms on E and the R_0-module of $n \times n$ hermitian matrices in R.

Remark. If we had assumed at the beginning that our automorphism $a \mapsto \bar{a}$ has period 2 or 1 (i.e. if we allow it to be the identity), then the results on bilinear and symmetric forms become special cases of the results of this section. However, the notational differences are sufficiently disturbing to warrant a repetition of the results as we have done.

Terminology

For some confusing reason, the group of automorphisms of a symmetric (resp. alternating, resp. hermitian) form on a vector space is called the **orthogonal** (resp. **symplectic**, resp. **unitary**) group of the form. The word orthogonal is especially unfortunate, because an orthogonal map preserves more than orthogonality: It also preserves the scalar product, i.e. length. Furthermore, the word symplectic is also unfortunate. It turns out that one can carry out a discussion of hermitian forms over certain division rings (having automorphisms of order 2), and their group of automorphisms have also been called symplectic, thereby creating genuine confusion with the use of the word relative to alternating forms.

In order to unify and improve the terminology, I have discussed the matter with several persons, and it seems that one could adopt the following conventions.

As said in the text, the group of automorphisms of any form f is denoted by $\mathrm{Aut}(f)$.

On the other hand, there is a standard form, described over the real numbers in terms of coordinates by

$$f(x, x) = x_1^2 + \cdots + x_n^2,$$

over the complex numbers by

$$f(x, x) = x_1\bar{x}_1 + \cdots + x_n\bar{x}_n,$$

and over the quaternions by the same formula as in the complex case. The group of automorphisms of this form would be called the **unitary group**, and be denoted by U_n. The points of this group in the reals (resp. complex, resp. quaternions) would be denoted by

$$U_n(\mathbf{R}), \qquad U_n(\mathbf{C}), \qquad U_n(\mathbf{K}),$$

and these three groups would be called the **real unitary group** (resp. **complex unitary group**, resp. **quaternion unitary group**). Similarly, the group of points of U_n in any subfield or subring k of the quaternions would be denoted by $U_n(k)$.

Finally, if f is the standard alternating form, whose matrix is

$$\begin{pmatrix} 0 & I_n \\ -I_n & 0 \end{pmatrix},$$

one would denote its group of automorphisms by A_{2n}, and call it the **alternating form group**, or simply the alternating group, if there is no danger of confusion with the permutation group. The group of points of the alternating form group in a field k would then be denoted by $A_{2n}(k)$.

As usual, the subgroup of $\mathrm{Aut}(f)$ consisting of those elements whose determinant is 1 would be denoted by adding the letter S in front, and would still be called the **special group**. In the four standard cases, this yields

$$SU_n(\mathbf{R}), \qquad SU_n(\mathbf{C}), \qquad SU_n(\mathbf{K}), \qquad SA_{2n}(k).$$

§8. THE SIMPLICITY OF $SL_2(F)/\pm 1$

Let F be a field. Let n be a positive integer. By $GL_n(F)$ we mean the group of $n \times n$ invertible matrices over F. By $SL_n(F)$ we mean the subgroup of those matrices whose determinant is equal to 1. By $PGL_n(F)$ we mean the factor group of $GL_n(F)$ by the subgroup of scalar matrices (which are in the center).

Similarly for $PSL_n(F)$. In this section, we are interested in giving an application of matrices to the group theoretic structure of SL_2. The analogous statements for SL_n with $n \geqq 3$ will be proved in the next section.

The **standard Borel subgroup** B of GL_2 is the group of all matrices

$$\begin{pmatrix} a & b \\ 0 & d \end{pmatrix}$$

with $a, b, d \in F$ and $ad \neq 0$. For the Borel subgroup of SL_2, we require in addition that $ad = 1$. By a **Borel subgroup** we mean a subgroup which is conjugate to the standard Borel subgroup (whether in GL_2 or SL_2). We let U be the group of matrices

$$u(b) = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}, \quad \text{with } b \in F.$$

We let A be the group of diagonal matrices

$$\begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}, \quad \text{with } a, d \in F^*.$$

Let

$$s(a) = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \quad \text{with } a \in F^*$$

and

$$w = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

For the rest of this section, we let

$$G = GL_2(F) \quad \text{or} \quad SL_2(F).$$

Lemma 8.1. *The matrices*

$$X(b) = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad Y(c) = \begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix}$$

generate $SL_2(F)$.

Proof. Multiplying an arbitrary element of $SL_2(F)$ by matrices of the above type on the right and on the left corresponds to elementary row and column operations, that is adding a scalar multiple of a row to the other, etc. Thus a given matrix can always be brought into a form

$$\begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}$$

by such multiplications. We want to express this matrix with $a \neq 1$ in the form

$$\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ b & 1 \end{pmatrix}\begin{pmatrix} 1 & c \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ d & 1 \end{pmatrix}.$$

Matrix multiplication will show that we can solve this equation, by selecting x arbitrarily $\neq 0$, then solving for b, c, and d successively so that

$$1 + bx = a, \quad c = \frac{-x}{1 + bx}, \quad d = \frac{-b}{1 + bc}.$$

Then one finds $1 + bc = (1 + xb)^{-1}$ and the two symmetric conditions

$$b + bcd + d = 0$$
$$c + bcx + x = 0,$$

so we get what we want, and thereby prove the lemma.

Let $\bar{U}$ be the group of lower matrices

$$\begin{pmatrix} 1 & 0 \\ c & 1 \end{pmatrix}.$$

Then we see that

$$wUw^{-1} = \bar{U}.$$

Also note the commutation relation

$$w\begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}w^{-1} = \begin{pmatrix} d & 0 \\ 0 & a \end{pmatrix},$$

so w normalizes A. Similarly,

$$wBw^{-1} = \bar{B}$$

is the group of lower triangular matrices.

We note that

$$B = AU = UA,$$

and also that A normalizes U.

There is a decomposition of G into disjoint subsets

$$G = B \cup BwB.$$

Indeed, view G as operating on the left of column vectors. The isotropy group of

$$e^1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

is obviously U. The orbit Be^1 consists of all column vectors whose second

component is 0. On the other hand,

$$we^1 = \begin{pmatrix} 0 \\ -1 \end{pmatrix},$$

and therefore the orbit Bwe^1 consists of all vectors whose second component is $\neq 0$, and whose first component is arbitrary. Since these two orbits of B and BwB cover the orbit Ge^1, it follows that the union of B and BwB is equal to G (because the isotropy group U is contained in B), and they are obviously disjoint. This decomposition is called the **Bruhat decomposition**.

Proposition 8.2. *The Borel subgroup B is a maximal proper subgroup.*

Proof. By the Bruhat decomposition, any element not in B lies in BwB, so the assertion follows since B, BwB cover G.

Theorem 8.3. *If F has at least four elements, then $SL_2(F)$ is equal to its own commutator group.*

Proof. We have the commutator relation (by matrix multiplication)

$$s(a)u(b)s(a)^{-1}u(b)^{-1} = u(ba^2 - b) = u(b(a^2 - 1)).$$

Let $G = SL_2(F)$ for this proof. We let G' be the commutator subgroup, and similarly let B' be the commutator subgroup of B. We prove the first assertion that $G = G'$. From the hypothesis that F has at least four elements, we can find an element $a \neq 0$ in F such that $a^2 \neq 1$, whence the commutator relation shows that $B' = U$. It follows that $G' \supset U$, and since G' is normal, we get

$$G' \supset wUw^{-1}.$$

From Lemma 8.1, we conclude that $G' = G$.

Let Z denote the center of G. It consists of $\pm I$, that is $\pm$ the identity 2×2 matrix if $G = SL_2(F)$; and Z is the subgroup of scalar matrices if $G = GL_2(F)$.

Theorem 8.4. *If F has at least four elements, then $SL_2(F)/Z$ is simple.*

The proof will result from two lemmas.

Lemma 8.5. *The intersection of all conjugates of B in G is equal to Z.*

Proof. We leave this to the reader, as a simple fact using conjugation with w.

Lemma 8.6. *Let $G = SL_2(F)$. If H is normal in G, then either $H \subset Z$ or $H \supset G'$.*

Proof. By the maximality of B we must have

$$HB = B \quad \text{or} \quad HB = G.$$

If $HB = B$ then $H \subset B$. Since H is normal, we conclude that H is contained in every conjugate of B, whence in the center by Lemma 8.5. On the other hand, suppose that $HB = G$. Write

$$w = hb$$

with $h \in H$ and $b \in B$. Then

$$wUw^{-1} = \bar{U} = hbUb^{-1}h^{-1} = hUh^{-1} \subset HU$$

because H is normal. Since $U \subset HU$ and $U, \bar{U}$ generate $SL_2(F)$, it follows that $HU = G$. Hence

$$G/H = HU/H \approx U/(U \cap H)$$

is abelian, whence $H \supset G'$, as was to be shown.

The simplicity of Theorem 8.4 is an immediate consequence of Lemma 8.6.

§9. THE GROUP $SL_n(F)$, $n \geqq 3$.

In this section we look at the case with $n \geqq 3$, and follow parts of Artin's *Geometric Algebra*, Chapter IV. (Artin even treats the case of a non-commutative division algebra as the group ring, but we omit this for simplicity.)

For $i, j = 1, \ldots, n$ and $i \neq j$ and $c \in F$, we let

$$E_{ij}(c) = \begin{pmatrix} 1 & & & \\ & 1 & & 0 \\ 0 & & \ddots & \\ & c_{ij} & & \\ 0 & & 0 & 1 \end{pmatrix}$$

be the matrix which differs from the unit matrix by having c in the ij-component instead of 0. We call such $E_{ij}(c)$ an **elementary matrix**. Note that

$$\det E_{ij}(c) = 1.$$

If A is any $n \times n$ matrix, then multiplication $E_{ij}(c)A$ on the left adds c times the j-th row to the i-th row of A. Multiplication $AE_{ij}(c)$ on the right adds c times the i-th column to the j-th column. We shall mostly multiply on the left.

For fixed $i \neq j$ the map

$$c \mapsto E_{ij}(c)$$

is a homomorphism of F into the multiplicative group of $n \times n$ matrices $GL_n(F)$.

Proposition 9.1. *The group $SL_n(F)$ is generated by the elementary matrices. If $A \in GL_n(F)$, then A can be written in the form*

$$A = SD,$$

where $S \in SL_n(F)$ and D is a diagonal matrix of the form

$$D = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & d \end{pmatrix}$$

so D has 1 on the diagonal except on the lower right corner, where the component is $d = \det(A)$.

Proof. Let $A \in GL_n(F)$. Since A is non-singular, the first component of some row is not zero, and by an elementary row operation, we can make $a_{11} \neq 0$. Adding a suitable multiple of the first row to the second row, we make $a_{21} \neq 0$, and then adding a suitable multiple of the second row to the first we make $a_{11} = 1$. Then we subtract multiples of the first row from the others to make $a_{i1} = 0$ for $i \neq 1$.

We now repeat the procedure with the second row and column, to make $a_{22} = 1$ and $a_{i2} = 0$ if $i > 2$. But then we can also make $a_{12} = 0$ by subtracting a suitable multiple of the second row from the first, so we can get $a_{i2} = 0$ for $i \neq 2$.

We repeat this procedure until we are stopped at $a_{nn} = d \neq 0$, and $a_{nj} = 0$ for $j \neq n$. Subtracting a suitable multiple of the last row from the preceding ones yields a matrix D of the form indicated in the statement of the theorem, and concludes the proof.

Theorem 9.2. *For $n \geqq 3$, $SL_n(F)$ is equal to its own commutator group.*

Proof. It suffices to prove that $E_{ij}(c)$ is a commutator. Using $n \geqq 3$, let $k \neq i, j$. Then by direct computation,

$$E_{ij}(c) = E_{ik}(c)E_{kj}(1)E_{ik}(-c)E_{kj}(-1)$$

expresses $E_{ij}(c)$ as a commutator. This proves the theorem.

We note that if a matrix M commutes with every element of $SL_n(F)$, then it must be a scalar matrix. Indeed, just the commutation with the elementary matrices

$$E_{ij}(1) = I + 1_{ij}$$

shows that M commutes with all matrices 1_{ij} (having 1 in the ij-component, 0 otherwise), so M commutes with all matrices, and is a scalar matrix. Taking the determinant shows that the center consists of $\boldsymbol{\mu}_n(F)I$, where $\boldsymbol{\mu}_n(F)$ is the group of n-th roots of unity in F.

We let Z be the center of $SL_n(F)$, so we have just seen that Z is the group of scalar matrices such that the scalar is an n-th root of unity. Then we define

$$PSL_n(F) = SL_n(F)/Z.$$

Theorem 9.3. *For $n \geqq 3$, $PSL_n(F)$ is simple.*

The rest of this section is devoted to the proof. We view $GL_n(F)$ as operating on the vector space $E = F^n$. If λ is a non-zero functional on E, we let

$$H_\lambda = \operatorname{Ker} \lambda,$$

and call H_λ (or simply H) the **hyperplane associated with** λ. Then $\dim H = n - 1$, and conversely, if H is a subspace of codimension 1, then E/H has dimension 1, and is the kernel of a functional.

An element $T \in GL_n(F)$ is called a **transvection** if it keeps every element of some hyperplane H fixed, and for all $x \in E$, we have

$$Tx = x + h \qquad \text{for some } h \in H.$$

Given any element $u \in H_\lambda$ we define a transvection T_u by

$$T_u x = x + \lambda(x)u.$$

Every transvection is of this type. If $u, v \in H_\lambda$, it is immediate that

$$T_{u+v} = T_u \circ T_v.$$

If T is a transvection and $A \in GL_n(F)$, then the conjugate ATA^{-1} is obviously a transvection.

The elementary matrices $E_{ij}(c)$ are transvections, and it will be useful to use them with this geometric interpretations, rather than formally as we did before. Indeed, let $e_1, \ldots, e_n$ be the standard unit vectors which form a basis of $F^{(n)}$. Then $E_{ij}(c)$ leaves e_k fixed if $k \neq j$, and the remaining vector e_j is moved by a multiple of e_i. We let H be the hyperplane generated by e_k with $k \neq j$, and thus see that $E_{ij}(c)$ is a transvection.

Lemma 9.4. *For $n \geqq 3$, the transvections $\neq I$ form a single conjugacy class in $SL_n(F)$.*

Proof. First, by picking a basis of a hyperplane $H = H_\lambda$ and using one more element to form a basis of $F^{(n)}$, one sees from the matrix of a transvection T that $\det T = 1$, i.e. transvections are in $SL_n(F)$.

Let T' be another transvection relative to a hyperplane H'. Say

$$Tx = x + \lambda(x)u \quad \text{and} \quad T'x = x + \lambda'(x)u'$$

with $u \in H$ and $u' \in H'$. Let z and z' be vectors such that $\lambda(z) = 1$ and $\lambda'(z') = 1$. Since a basis for H together with z is a basis for $F^{(n)}$, and similarly a basis for H' together with z' is a basis for $F^{(n)}$, there exists an element $A \in GL_n(F)$ such that

$$Au = u', \qquad AH = H', \qquad Az = z'.$$

It is then immediately verified that

$$ATA^{-1} = T',$$

so T, T' are conjugate in $GL_n(F)$. But in fact, using $n \geqq 3$, the hyperplanes H, H' contain vectors which are independent. We can change the image of a basis vector in H' which is independent of u' by some factor in F so as to make $\det A = 1$, so $A \in SL_n(F)$. This proves the lemma.

We now want to show that certain subgroups of $GL_n(F)$ are either contained in the center, or contain $SL_n(F)$. Let G be a subgroup of $GL_n(F)$. We say that G is **SL_n-invariant** if

$$AGA^{-1} \subset G \quad \text{for all } A \in SL_n(F).$$

Lemma 9.5. *Let $n \geqq 3$. Let G be SL_n-invariant, and suppose that G contains a transvection $T \neq I$. Then $SL_n(F) \subset G$.*

Proof. By Lemma 9.4, all transvections are conjugate, and the set of transvections contains the elementary matrices which generate $SL_n(F)$ by Proposition 9.1, so the lemma follows.

Theorem 9.6. *Let $n \geqq 3$. If G is a subgroup of $GL_n(F)$ which is SL_n-invariant and which is not contained in the center of $GL_n(F)$, then $SL_n(F) \subset G$.*

Proof. By the preceding lemma, it suffices to prove that G contains a transvection, and this is the key step in the proof of Theorem 9.3.

We start with an element $A \in G$ which moves some line. This is possible since G is not contained in the center. So there exists a vector $u \neq 0$ such that Au is not a scalar multiple of u, say $Au = v$. Then u, v are contained in some hyperplane $H = \operatorname{Ker} \lambda$. Let $T = T_u$ and let

$$B = ATA^{-1}T^{-1}.$$

Then

$$ATA^{-1} \neq T \quad \text{and} \quad B = ATA^{-1}T^{-1} \neq I.$$

This is easily seen by applying say B to an arbitrary vector x, and using the definition of T_u. In each case, for some x the left-hand side cannot equal the right-hand side.

For any vector $x \in F^{(n)}$ we have

$$Bx - x \in (u, v),$$

where (u, v) is the plane generated by u, v. It follows that $BH \subset H$, so

$$BH = H \quad \text{and} \quad Bx - x \in H.$$

We now distinguish two cases to conclude the proof. First assume that B commutes with all transvections with respect to H. Let $w \in H$. Then from the definitions, we find for any vector x:

$$BT_w x = Bx + \lambda(x)Bw$$

$$T_w Bx = Bx + \lambda(Bx)w = Bx + \lambda(x)w.$$

Since we are in the case $BT_w = T_w B$, it follows that $Bw = w$. Therefore B leaves every vector of H fixed. Since we have seen that $Bx - x \in H$ for all x, it follows that B is a transvection and is in G, thus proving the theorem in this case.

Second, suppose there is a transvection T_w with $w \in H$ such that B does not commute with T_w. Let

$$C = BT_w B^{-1} T_w^{-1}.$$

Then $C \neq I$ and $C \in G$. Furthermore C is a product of T_w^{-1} and $BT_w B^{-1}$ whose hyperplanes are H and BH, which is also H by what we have already proved. Therefore C is a transvection, since it is a product of transvections with the same hyperplane. And $C \in G$. This concludes the proof in the second case, and also concludes the proof of Theorem 9.6.

We now return to the main theorem, that $PSL_n(F)$ is simple. Let $\bar{G}$ be a normal subgroup of $PSL_n(F)$, and let G be its inverse image in $SL_n(F)$. Then G is SL_n-invariant, and if $\bar{G} \neq 1$, then G is not equal to the center of $SL_n(F)$. Therefore G contains $SL_n(F)$ by Theorem 9.6, and therefore $\bar{G} = PSL_n(F)$, thus proving that $PSL_n(F)$ is simple.

Example. By Exercise 41 of Chapter I, or whatever other means, one sees that $PSL_2(\mathbf{F}_5) \approx A_5$ (where $\mathbf{F}_5$ is the finite field with 5 elements). While you are in the mood, show also that

$$PGL_2(\mathbf{F}_3) \approx S_4 \quad \text{but} \quad SL_2(\mathbf{F}_3) \not\approx S_4; \quad PSL_2(\mathbf{F}_3) \approx A_4.$$

EXERCISES

1. Interpret the rank of a matrix A in terms of the dimensions of the image and kernel of the linear map L_A.

2. (a) Let A be an invertible matrix in a commutative ring R. Show that $({}^tA)^{-1} = {}^t(A^{-1})$.
   (b) Let f be a non-singular bilinear form on the module E over R. Let A be an R-automorphism of E. Show that $({}^tA)^{-1} = {}^t(A^{-1})$. Prove the same thing in the hermitian case, i.e. $(A^*)^{-1} = (A^{-1})^*$.

3. Let V, W be finite dimensional vector spaces over a field k. Suppose given non-degenerate bilinear forms on V and W respectively, denoted both by $\langle\ ,\ \rangle$. Let $L: V \to W$ be a surjective linear map and let ${}^tL$ be its transpose; that is, $\langle Lv, w\rangle = \langle v, {}^tLw\rangle$ for $v \in V$ and $w \in W$.
   (a) Show that ${}^tL$ is injective.
   (b) Assume in addition that if $v \in V$, $v \neq 0$ then $\langle v, v\rangle \neq 0$. Show that
   $$V = \operatorname{Ker} L \oplus \operatorname{Im} {}^tL,$$
   and that the two summands are orthogonal. (Cf. Exercise 33 for an example.)

4. Let $A_1, \ldots, A_r$ be row vectors of dimension n, over a field k. Let $X = (x_1, \ldots, x_n)$. Let $b_1, \ldots, b_r \in k$. By a system of linear equations in k one means a system of type
   $$A_1 \cdot X = b_1, \ldots, A_r \cdot X = b_r.$$
   If $b_1 = \cdots = b_r = 0$, one says the system is homogeneous. We call n the number of variables, and r the number of equations. A solution X of the homogeneous system is called **trivial** if $x_i = 0$, $i = 1, \ldots, n$.
   (a) Show that a homogeneous system of r linear equations in n unknowns with $n > r$ always has a non-trivial solution.
   (b) Let L be a system of homogeneous linear equations over a field k. Let k be a subfield of k'. If L has a non-trivial solution in k', show that it has a non-trivial solution in k.

5. Let M be an $n \times n$ matrix over a field k. Assume that $\operatorname{tr}(MX) = 0$ for all $n \times n$ matrices X in k. Show that $M = O$.

6. Let S be a set of $n \times n$ matrices over a field k. Show that there exists a column vector $X \neq 0$ of dimension n in k, such that $MX = X$ for all $M \in S$ if and only if there exists such a vector in some extension field k' of k.

7. Let $\mathbf{H}$ be the division ring over the reals generated by elements i, j, k such that $i^2 = j^2 = k^2 = -1$, and
   $$ij = -ji = k, \qquad jk = -kj = i, \qquad ki = -ik = j.$$
   Then $\mathbf{H}$ has an automorphism of order 2, given by
   $$a_0 + a_1 i + a_2 j + a_3 k \mapsto a_0 - a_1 i - a_2 j - a_3 k.$$
   Denote this automorphism by $\alpha \mapsto \bar{\alpha}$. What is $\alpha\bar{\alpha}$? Show that the theory of hermitian

forms can be carried out over **H**, which is called the division ring of **quaternions** (or by abuse of language, the non-commutative field of quaternions).

8. Let N be a strictly upper triangular $n \times n$ matrix, that is $N = (a_{ij})$ and $a_{ij} = 0$ if $i \geqq j$. Show that $N^n = 0$.

9. Let E be a vector space over k, of dimension n. Let $T: E \to E$ be a linear map such that T is nilpotent, that is $T^m = 0$ for some positive integer m. Show that there exists a basis of E over k such that the matrix of T with respect to this basis is strictly upper triangular.

10. If N is a nilpotent $n \times n$ matrix, show that $I + N$ is invertible.

11. Let R be the set of all upper triangular $n \times n$ matrices (a_{ij}) with a_{ij} in some field k, so $a_{ij} = 0$ if $i > j$. Let J be the set of all strictly upper triangular matrices. Show that J is a two-sided ideal in R. How would you describe the factor ring R/J?

12. Let G be the group of upper triangular matrices with non-zero diagonal elements. Let H be the subgroup consisting of those matrices whose diagonal element is 1. (Actually prove that H is a subgroup). How would you describe the factor group G/H?

13. Let R be the ring of $n \times n$ matrices over a field k. Let L be the subset of matrices which are 0 except on the first column.
    (a) Show that L is a left ideal.
    (b) Show that L is a minimal left ideal; that is, if $L' \subset L$ is a left ideal and $L' \neq 0$, then $L' = L$. (For more on this situation, see Chapter VII, §5.)

14. Let F be any field. Let D be the subgroup of diagonal matrices in $GL_n(F)$. Let N be the normalizer of D in $GL_n(F)$. Show that N/D is isomorphic to the symmetric group on n elements.

15. Let F be a finite field with q elements. Show that the order of $GL_n(F)$ is

$$(q^n - 1)(q^n - q) \cdots (q^n - q^{n-1}) = q^{n(n-1)/2} \prod_{i=1}^{n} (q^i - 1).$$

[*Hint*: Let $x_1, \dots, x_n$ be a basis of F^n. Any element of $GL_n(F)$ is uniquely determined by its effect on this basis, and thus the order of $GL_n(F)$ is equal to the number of all possible bases. If $A \in GL_n(F)$, let $Ax_i = y_i$. For y_1 we can select any of the $q^n - 1$ non-zero vectors in F^n. Suppose inductively that we have already chosen $y_1, \dots, y_r$ with $r < n$. These vectors span a subspace of dimension r which contains q^r elements. For y_{i+1} we can select any of the $q^n - q^r$ elements outside of this subspace. The formula drops out.]

16. Again let F be a finite field with q elements. Show that the order of $SL_n(F)$ is

$$q^{n(n-1)/2} \prod_{i=2}^{n} (q^i - 1);$$

and that the order of $PSL_n(F)$ is

$$\frac{1}{d} q^{n(n-1)/2} \prod_{i=2}^{n-1} (q^i - 1),$$

where d is the greatest common divisor of n and $q - 1$.

17. Let F be a finite field with q elements. Show that the group of all upper triangular matrices with 1 on the diagonal is a Sylow subgroup of $GL_n(F)$ and of $SL_n(F)$.

18. The reduction map $\mathbf{Z} \to \mathbf{Z}/N\mathbf{Z}$, where N is a positive integer defines a homomorphism

$$SL_2(\mathbf{Z}) \to SL_2(\mathbf{Z}/N\mathbf{Z}).$$

Show that this homomorphism is surjective. [*Hint*: Use elementary divisors, i.e. the structure of submodules of rank 2 over the principal ring $\mathbf{Z}$.]

19. Show that the order of $SL_2(\mathbf{Z}/N\mathbf{Z})$ is equal to

$$N^3 \prod_{p \mid N} \left(1 - \frac{1}{p^2}\right),$$

where the product is taken over all primes dividing N.

20. Show that one has an exact sequence

$$1 \to SL_2(\mathbf{Z}/N\mathbf{Z}) \to GL_2(\mathbf{Z}/N\mathbf{Z}) \xrightarrow{\det} (\mathbf{Z}/N\mathbf{Z})^* \to 1.$$

In fact, show that

$$GL_2(\mathbf{Z}/N\mathbf{Z}) = SL_2(\mathbf{Z}/N\mathbf{Z})G_N,$$

where G_N is the group of matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & d \end{pmatrix} \quad \text{with} \quad d \in (\mathbf{Z}/N\mathbf{Z})^*.$$

21. Show that $SL_2(\mathbf{Z})$ is generated by the matrices

$$\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

22. Let p be a prime $\geqq 5$. Let G be a subgroup of $SL_2(\mathbf{Z}/p^n\mathbf{Z})$ with $n \geqq 1$. Assume that the image of G in $SL_2(\mathbf{Z}/p\mathbf{Z})$ under the natural homomorphism is all of $SL_2(\mathbf{Z}/p\mathbf{Z})$. Prove that $G = SL_2(\mathbf{Z}/p^n\mathbf{Z})$.

Note. Exercise 22 is a generalization by Serre of a result of Shimura; see Serre's *Abelian ℓ-adic Representations and elliptic curves*, Benjamin, 1968, IV, §3, Lemma 3. See also my exposition in *Elliptic Functions*, Springer Verlag, reprinted from Addison-Wesley, 1973, Chapter 17, §4.

23. Let k be a field in which every quadratic polynomial has a root. Let B be the Borel subgroup of $GL_2(k)$. Show that G is the union of all the conjugates of B. (This cannot happen for finite groups!)

24. Let A, B be square matrices of the same size over a field k. Assume that B is non-singular. If t is a variable, show that $\det(A + tB)$ is a polynomial in t, whose leading coefficient is $\det(B)$, and whose constant term is $\det(A)$.

25. Let $a_{11}, \ldots, a_{1n}$ be elements from a principal ideal ring, and assume that they generate the unit ideal. Suppose $n > 1$. Show that there exists a matrix (a_{ij}) with this given first row, and whose determinant is equal to 1.

26. Let A be a commutative ring, and $I = (x_1, \dots, x_r)$ an ideal. Let $c_{ij} \in A$ and let

$$y_i = \sum_{j=1}^{r} c_{ij} x_j.$$

Let $I' = (y_1, \dots, y_r)$. Let $D = \det(c_{ij})$. Show that $DI \subset I'$.

27. Let L be a free module over $\mathbf{Z}$ with basis $e_1, \dots, e_n$. Let M be a free submodule of the same rank, with basis $u_1, \dots, u_n$. Let $u_i = \sum c_{ij} e_j$. Show that the index $(L : M)$ is given by the determinant:

$$(L : M) = |\det(c_{ij})|.$$

28. **(The Dedekind determinant).** Let G be a finite commutative group and let F be the vector space of functions of G into $\mathbf{C}$. Show that the characters of G (homomorphisms of G into the roots of unity) form a basis for this space. If $f : G \to \mathbf{C}$ is a function, show that for $a, b \in G$.

$$\det(f(ab^{-1})) = \prod_{\chi} \sum_{a \in G} \chi(a) f(a),$$

where the product is taken over all characters. [*Hint*: Use both the characters and the charąacteristic functions of elements of G as bases for F, and consider the linear map

$$T = \sum f(a) T_a,$$

where T_a is translation by a.] Also show that

$$\det(f(ab^{-1})) = \left(\sum_{a \in G} f(a)\right) \det(f(ab^{-1}) - f(b^{-1})),$$

where the determinant on the left is taken for all $a, b \in G$, and the determinant on the right is taken only for $a, b \neq 1$.

29. Let $\mathfrak{g}$ be a module over the commutative ring R. A bilinear map $\mathfrak{g} \times \mathfrak{g} \to \mathfrak{g}$, written $(x, y) \mapsto [x, y]$, is said to make $\mathfrak{g}$ a **Lie algebra** if $[x, x] = 0$ and

$$[[x, y], z] + [[y, z], x] + [[z, x], y] = 0$$

for all $x, y, z \in \mathfrak{g}$.

(a) Let $M_n(R)$ be the ring of matrices over R. If $x, y \in M_n(R)$, show that the product

$$(x, y) \mapsto [x, y] = xy - yx$$

makes $M_n(R)$ into a Lie algebra.

(b) Let $\mathfrak{g}$ be a Lie algebra. Let $x \in \mathfrak{g}$, and let L_x, $L(x)$ or Lie x be the linear map given by $L_x(y) = [x, y]$. Show that L_x is a derivation of $\mathfrak{g}$ into itself (i.e. satisfies the rule $D([y, z]) = [Dy, z] + [y, Dz]$).

(c) Show that the map $x \mapsto L_x$ is a Lie homomorphism of $\mathfrak{g}$ into the module of derivations of $\mathfrak{g}$ into itself.

30. Given a set of polynomials $\{P_\nu(X_{ij})\}$ in the polynomial ring $R[X_{ij}]$ $(1 \leqq i, j \leqq n)$, a zero of this set in R is a matrix $x = (x_{ij})$ such that $x_{ij} \in R$ and $P_\nu(x_{ij}) = 0$ for all ν. We use vector notation, and write $(X) = (X_{ij})$. We let $G(R)$ denote the set of zeros

of our set of polynomials $\{P_\nu\}$. Thus $G(R) \subset M_n(R)$, and if R' is any commutative associative R-algebra we have $G(R') \subset M_n(R')$. We say that the set $\{P_\nu\}$ defines an **algebraic group over** R if $G(R')$ is a subgroup of the group $GL_n(R')$ for all R' (where $GL_n(R')$ is the multiplicative group of invertible matrices in R').

As an example, the group of matrices satisfying the equation ${}^tXX = I_n$ is an algebraic group.

Let R' be the R-algebra which is free, with a basis $\{1, t\}$ such that $t^2 = 0$. Thus $R' = R[t]$. Let $\mathfrak{g}$ be the set of matrices $x \in M_n(R)$ such that $I_n + tx \in G(R[t])$. Show that $\mathfrak{g}$ is a Lie algebra. [*Hint*: Note that

$$P_\nu(I_n + tX) = P_\nu(I_n) + \text{grad } P_\nu(I_n)tX.$$

Use the algebra $R[t, u]$ where $t^2 = u^2 = 0$ to show that if $I_n + tx \in G(R[t])$ and $I_n + uy \in G(R[u])$ then $[x, y] \in \mathfrak{g}$.]

(I have taken the above from the first four pages of [Se 65]. For more information on Lie algebras and Lie Groups, see [Bo 82] and [Ja 79].

[Bo 82] N. BOURBAKI, *Lie Algebras and Lie Groups*, Masson, 1982

[Ja 79] N. JACOBSON, *Lie Algebras*, Dover, 1979 (reprinted from Interscience, 1962)

[Se 65] J. P. SERRE, *Lie Algebras and Lie Groups*, Benjamin, 1965. Reprinted Springer Lecture Notes 1500. Springer/Verlag 1992

Non-commutative cocycles

Let K be a finite Galois extension of a field k. Let $\Gamma = GL_n(K)$, and $G = \text{Gal}(K/k)$. Then G operates on Γ. By a **cocycle** of G in Γ we mean a family of elements $\{A(\sigma)\}$ satisfying the relation

$$A(\sigma)\sigma A(\tau) = A(\sigma\tau).$$

We say that the cocycle **splits** if there exists $B \in \Gamma$ such that

$$A(\sigma) = B^{-1}\sigma B \qquad \text{for all } \sigma \in G.$$

In this non-commutative case, cocycles do not form a group, but one could define an equivalence relation to define cohomology classes. For our purposes here, we care only whether a cocycle splits or not. When every cocycle splits, we also say that $H^1(G, \Gamma) = 0$ (or 1).

31. Prove that $H^1(G, GL_n(K)) = 1$. [*Hint*: Let $\{e_1, \ldots, e_N\}$ be a basis of $\text{Mat}_n(k)$ over k, say the matrices with 1 in some component and 0 elsewhere. Let

$$x = \sum_{i=1}^{N} x_i e_i$$

with variables x_i. There exists a polynomial $P(X)$ such that x is invertible if and only if $P(x_1, \ldots, x_N) \neq 0$. Instead of $P(x_1, \ldots, x_N)$ we also write $P(x)$. Let $\{A(\sigma)\}$ be a cocycle. Let $\{t_\sigma\}$ be algebraically independent variables over k. Then

$$P\left(\sum_{\gamma \in G} t_\gamma A(\gamma)\right) \neq 0$$

because the polynomial does not vanish when one t_γ is replaced by 1 and the others are replaced by 0. By the algebraic independence of automorphisms from Galois theory, there exists an element $y \in K$ such that if we put

$$B = \sum_\gamma (\gamma y) A(\gamma)$$

then $P(B) \neq 0$, so B is invertible. It is then immediately verified that $A(\sigma) = B\sigma B^{-1}$. But when k is finite, cf. my *Algebraic Groups over Finite Fields*, Am. J. Vol 78 No. 3, 1956.]

32. **Invariant bases.** (Kolchin-Lang, *Proc. AMS* Vol 11 No. 1, 1960). Let K be a finite Galois extension of k, $G = \text{Gal}(K/k)$ as in the preceding exercise. Let V be a finite-dimensional vector space over K, and suppose G operates on V in such a way that $\sigma(av) = \sigma(a)\sigma(v)$ for $a \in K$ and $v \in V$. Prove that there exists a basis $\{w_1, \ldots, w_n\}$ such that $\sigma w_i = w_i$ for all $i = 1, \ldots, n$ and all $\sigma \in G$ (an invariant basis). *Hint*: Let $\{v_1, \ldots, v_n\}$ be any basis, and let

$$\sigma \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = A(\sigma) \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

where $A(\sigma)$ is a matrix in $GL_n(K)$. Solve for B in the equation $(\sigma B)A(\sigma) = B$, and let

$$\begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} = B \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}.$$

The next exercises on harmonic polynomials have their source in Whittaker, *Math. Ann.* 1902; see also Whittaker and Watson, *Modern Analysis*, Chapter XIII.

33. **Harmonic polynomials.** Let $\text{Pol}(n, d)$ denote the vector space of homogeneous polynomials of degree d in n variables $X_1, \ldots, X_n$ over a field k of characteristic 0. For an n-tuple of integers $(\nu_1, \ldots, \nu_n)$ with $\nu_i \geqq 0$ we denote by $M_{(\nu)}$ as usual the monomial

$$M_{(\nu)}(X) = X_1^{\nu_1} \cdots X_n^{\nu_n}.$$

Prove:

(a) The number of monomials of degree d is $\binom{n-1+d}{n-1}$, so this number is the dimension of $\text{Pol}(n, d)$.

(b) Let $(D) = (D_1, \ldots, D_n)$ where D_i is the partial derivative with respect to the i-th variable. Then we can define $P(D)$ as usual. For $P, Q \in \text{Pol}(n, d)$, define

$$\langle P, Q \rangle = P(D)Q(0).$$

Prove that this defines a symmetric non-degenerate scalar product on $\text{Pol}(n, d)$. If k is not real, it may happen that $P \neq 0$ but $\langle P, P \rangle = 0$. However, if the ground field is real, then $\langle P, P \rangle > 0$ for $P \neq 0$. Show also that the monomials of degree d form an orthogonal basis. What is $\langle M_{(\nu)}, M_{(\nu)} \rangle$?

(c) The map $P \mapsto P(D)$ is an isomorphism of $\text{Pol}(n, d)$ onto its dual.

(d) Let $\Delta = D_1^2 + \cdots + D_n^2$. Note that Δ: Pol$(n, d) \to$ Pol$(n, d-2)$ is a linear map. Prove that Δ is surjective.

(e) Define Har$(n, d) =$ KerΔ = vector space of **harmonic homogeneous polynomials** of degree d. Prove that

$$\dim \text{Har}(n, d) = (n + d - 3)!(n + 2d - 2)/(n - 2)!d!.$$

In particular, if $n = 3$, then dim Har$(3, d) = 2d + 1$.

(f) Let $r^2 = X_1^2 + \cdots + X_n^2$. Let S denote multiplication by r^2. Show that

$$\langle \Delta P, Q\rangle = \langle P, SQ\rangle \text{ for } P \in \text{Pol}(n, d) \text{ and } Q \in \text{Pol}(n, d-2),$$

so ${}^t\Delta = S$. More generally, for $R \in$ Pol(n, m) and $Q \in$ Pol$(n, d - m)$ we have

$$\langle R(D)P, Q\rangle = \langle P, RQ\rangle.$$

(g) Show that $[\Delta, S] = 4d + 2n$ on Pol(n, d). Here $[\Delta, S] = \Delta \circ S - S \circ \Delta$. Actually, $[\Delta, S] = 4E + 2n$, where E is the Euler operator $E = \sum X_i D_i$, which is, however, the degree operator on homogeneous polynomials.

(h) Prove that Pol$(n, d) =$ Har$(n, d) \oplus r^2$Pol$(n, d - 2)$ and that the two summands are orthogonal. This is a classical theorem used in the theory of the Laplace operator.

(i) Let $(c_1, \ldots, c_n) \in k^n$ be such that $\sum c_i^2 = 0$. Let

$$H_c^d(X) = (c_1X_1 + \cdots + c_nX_n)^d.$$

Show that H_c^d is harmonic, i.e. lies in Har(n, d).

(j) For any $Q \in$ Pol(n, d), and a positive integer m, show that

$$Q(D)H_c^m(X) = m(m-1)\cdots(m-d+1)Q(c)H_c^{m-d}(X).$$

34. (Continuation of Exercise 33). Prove:

Theorem. *Let k be algebraically closed of characteristic* 0. *Let $n \geqq 3$. Then* Har(n, d) *as a vector space over k is generated by all polynomials H_c^d with $(c) \in k^n$ such that $\sum c_i^2 = 0$.*

[*Hint*: Let $Q \in$ Har(n, d) be orthogonal to all polynomials H_c^d with $(c) \in k^n$. By Exercise 33(h), it suffices to prove that $r^2 | Q$. But if $\sum c_i^2 = 0$, then by Exercise 33(j) we conclude that $Q(c) = 0$. By the Hilbert Nullstellensatz, it follows that there exists a polynomial $F(X)$ such that

$$Q(X)^s = r^2(X)F(X) \text{ for some positive integer } s.$$

But $n \geqq 3$ implies that $r^2(X)$ is irreducible, so $r^2(X)$ divides $Q(X)$.]

35. (Continuation of Exercise 34). Prove that the representation of $O(n) = U_n(\mathbf{R})$ on Har(n, d) is irreducible.
Readers will find a proof in the following:

S. Helgason, *Topics in Harmonic Analysis on Homogeneous Spaces*, Birkhäuser, 1981 (see especially §3, Theorem 3.1(ii))

N. Vilenkin, *Special Functions and the Theory of Group Representations*, AMS Translations of mathematical monographs **Vol. 22**, 1968 (Russian original, 1965), Chapter IX, §2.

R. HOWE and E. C. TAN, *Non-Abelian Harmonic Analysis*, Universitext, Springer Verlag, New York, 1992.

The Howe-Tan proof runs as follows. We now use the hermitian product

$$\langle P, Q \rangle = \int_{\mathbf{S}^{n-1}} P(x)\,\overline{Q(x)}\,d\sigma(x),$$

where σ is the rotation invariant measure on the $(n-1)$-sphere $\mathbf{S}^{n-1}$. Let $e_1, \ldots, e_n$ be the unit vectors in $\mathbf{R}^n$. We can identify $O(n-1)$ as the subgroup of $O(n)$ leaving e_n fixed. Observe that $O(n)$ operates on $\mathrm{Har}(n, d)$, say on the right by composition $P \mapsto P \circ A$, $A \in O(n)$, and this operation commutes with Δ. Let

$$\lambda\colon \mathrm{Har}(n, d) \to \mathbf{C}$$

be the functional such that $\lambda(P) = P(e_n)$. Then λ is $O(n-1)$-invariant, and since the hermitian product is non-degenerate, there exists a harmonic polynomial Q_n such that

$$\lambda(P) = \langle P, Q_n \rangle \quad \text{for all } P \in \mathrm{Har}(n, d).$$

Let $M \subset \mathrm{Har}(n, d)$ be an $O(n)$-submodule. Then the restriction λ_M of λ to M is nontrivial because $O(n)$ acts transitively on $\mathbf{S}^{n-1}$. Let Q_n^M be the orthogonal projection of Q_n on M. Then Q_n^M is $O(n-1)$-invariant, and so is a linear combination

$$Q_n^M(x) = \sum_{j+2k=d} c_j\, x_n^j\, r_{n-1}^{2k}.$$

Furthermore Q_n^H is harmonic. From this you can show that Q_n^H is uniquely determined, by showing the existence of recursive relations among the coefficients c_j. Thus the submodule M is uniquely determined, and must be all of $\mathrm{Har}(n, d)$.

Irreducibility of $\mathfrak{sl}_n(F)$.

36. Let F be a field of characteristic 0. Let $\mathfrak{g} = \mathfrak{sl}_n(F)$ be the vector space of matrices with trace 0, with its Lie algebra structure $[X, Y] = XY - YX$. Let E_{ij} be the matrix having (i, j)-component 1 and all other components 0. Let $G = SL_n(F)$. Let A be the multiplicative group of diagonal matrices over F.
    (a) Let $H_i = E_{ii} - E_{i+1,i+1}$ for $i = 1, \ldots, n-1$. Show that the elements E_{ij} $(i \neq j)$, $H_1, \ldots, H_{n-1}$ form a basis of $\mathfrak{g}$ over F.
    (b) For $g \in G$ let $\mathbf{c}(g)$ be the conjugation action on $\mathfrak{g}$, that is $\mathbf{c}(g)X = gXg^{-1}$. Show that each E_{ij} is an eigenvector for this action restricted to the group A.
    (c) Show that the conjugation representation of G on $\mathfrak{g}$ is irreducible, that is, if $V \neq 0$ is a subspace of $\mathfrak{g}$ which is $\mathbf{c}(G)$-stable, then $V = \mathfrak{g}$. *Hint:* Look up the sketch of the proof in [JoL 01], Chapter VII, Theorem 1.5, and put in all the details. Note that for $i \neq j$ the matrix E_{ij} is nilpotent, so for variable t, the exponential series $\exp(tE_{ij})$ is actually a polynomial. The derivative with respect to t can be taken in the formal power series $F[[t]]$, not using limits. If X is a matrix, and $x(t) = \exp(tX)$, show that

$$\left.\frac{d}{dt} x(t)\,Y x(t)^{-1}\right|_{t=0} = XY - YX = [X, Y].$$

CHAPTER XIV

Representation of One Endomorphism

We deal here with one endomorphism of a module, actually a free module, and especially a finite dimensional vector space over a field k. We obtain the Jordan canonical form for a representing matrix, which has a particularly simple shape when k is algebraically closed. This leads to a discussion of eigenvalues and the characteristic polynomial. The main theorem can be viewed as giving an example for the general structure theorem of modules over a principal ring. In the present case, the principal ring is the polynomial ring $k[X]$ in one variable.

§1. REPRESENTATIONS

Let k be a commutative ring and E a module over k. As usual, we denote by $\text{End}_k(E)$ the ring of k-endomorphisms of E, i.e. the ring of k-linear maps of E into itself.

Let R be a k-algebra (given by a ring-homomorphism $k \to R$ which allows us to consider R as a k-module). By a **representation** of R in E one means a k-algebra homomorphism $R \to \text{End}_k(E)$, that is a ring-homomorphism

$$\rho : R \to \text{End}_k(E)$$

which makes the following diagram commutative:

$$\begin{array}{ccc} R & \longrightarrow & \text{End}_k(E) \\ & \nwarrow \quad \nearrow & \\ & k & \end{array}$$

[As usual, we view $\text{End}_k(E)$ as a k-algebra; if I denotes the identity map of E, we have the homomorphism of k into $\text{End}_k(E)$ given by $a \mapsto aI$. We shall also use I to denote the unit matrix if bases have been chosen. The context will always make our meaning clear.]

We shall meet several examples of representations in the sequel, with various types of rings (both commutative and non-commutative). In this chapter, the rings will be commutative.

We observe that E may be viewed as an $\text{End}_k(E)$ module. Hence E may be viewed as an R-module, defining the operation of R on E by letting

$$(x, v) \mapsto \rho(x)v$$

for $x \in R$ and $v \in E$. We usually write xv instead of $\rho(x)v$.

A subgroup F of E such that $RF \subset F$ will be said to be an **invariant** submodule of E. (It is both R-invariant and k-invariant.) We also say that it is invariant under the representation.

We say that the representation is **irreducible**, or **simple**, if $E \neq 0$, and if the only invariant submodules are 0 and E itself.

The purpose of representation theories is to determine the structure of all representations of various interesting rings, and to classify their irreducible representations. In most cases, we take k to be a field, which may or may not be algebraically closed. The difficulties in proving theorems about representations may therefore lie in the complication of the ring R, or the complication of the field k, or the complication of the module E, or all three.

A representation ρ as above is said to be **completely reducible** or **semi-simple** if E is an R-direct sum of R-submodules E_i,

$$E = E_1 \oplus \cdots \oplus E_m$$

such that each E_i is irreducible. We also say that E is completely reducible. It is not true that all representations are completely reducible, and in fact those considered in this chapter will not be in general. Certain types of completely reducible representations will be studied later.

There is a special type of representation which will occur very frequently. Let $v \in E$ and assume that $E = Rv$. We shall also write $E = (v)$. We then say that E is **principal** (over R), and that the representation is **principal**. If that is the case, the set of elements $x \in R$ such that $xv = 0$ is a left ideal $\mathfrak{a}$ of R (obvious). The map of R onto E given by

$$x \mapsto xv$$

induces an isomorphism of R-modules,

$$R/\mathfrak{a} \to E$$

(viewing R as a left module over itself, and $R/\mathfrak{a}$ as the factor module). In this map, the unit element 1 of R corresponds to the generator v of E.

As a matter of notation, if $v_1, \dots, v_n \in E$, we let $(v_1, \dots, v_n)$ denote the submodule of E generated by $v_1, \dots, v_n$.

Assume that E has a decomposition into a direct sum of R-submodules

$$E = E_1 \oplus \cdots \oplus E_s.$$

Assume that each E_i is free and of dimension $\geqq 1$ over k. Let $\mathcal{B}_1, \dots, \mathcal{B}_s$ be bases for $E_1, \dots, E_s$ respectively over k. Then $\{\mathcal{B}_1, \dots, \mathcal{B}_s\}$ is a basis for E. Let $\varphi \in R$, and let φ_i be the endomorphism induced by φ on E_i. Let M_i be the matrix of φ_i with respect to the basis $\mathcal{B}_i$. Then the matrix M of φ with respect to $\{\mathcal{B}_1, \dots, \mathcal{B}_s\}$ looks like

$$\begin{pmatrix} M_1 & 0 & \cdots & 0 \\ 0 & M_2 & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & 0 \\ 0 & \cdots & 0 & M_s \end{pmatrix}.$$

A matrix of this type is said to be decomposed into **blocks,** $M_1, \dots M_s$. When we have such a decomposition, the study of φ or its matrix is completely reduced (so to speak) to the study of the blocks.

It does not always happen that we have such a reduction, but frequently something almost as good happens. Let E' be a submodule of E, invariant under R. Assume that there exists a basis of E' over k, say $\{v_1, \dots, v_m\}$, and that this basis can be completed to a basis of E,

$$\{v_1, \dots, v_m, v_{m+1}, \dots, v_n\}.$$

This is always the case if k is a field.

Let $\varphi \in R$. Then the matrix of φ with respect to this basis has the form

$$\begin{pmatrix} M' & * \\ 0 & M'' \end{pmatrix}.$$

Indeed, since E' is mapped into itself by φ, it is clear that we get M' in the upper left, and a zero matrix below it. Furthermore, for each $j = m + 1, \dots, n$ we can write

$$\varphi v = c_{j1}v_1 + \dots + c_{jm}v_m + c_{j,m+1}v_{m+1} + \dots + c_{jn}v_n.$$

The transpose of the matrix (c_{ji}) then becomes the matrix

$$\begin{pmatrix} * \\ M'' \end{pmatrix}$$

occurring on the right in the matrix representing φ.

Furthermore, consider an exact sequence

$$0 \to E' \to E \to E'' \to 0.$$

Let $\bar{v}_{m+1}, \dots, \bar{v}_n$ be the images of $v_{m+1}, \dots, v_n$ under the canonical map $E \to E''$. We can define a linear map

$$\varphi'' : E'' \to E''$$

in a natural way so that $(\overline{\varphi v}) = \varphi''(\bar{v})$ for all $v \in E$. Then it is clear that the matrix of φ'' with respect to the basis $\{\bar{v}_1, \dots, \bar{v}_n\}$ is M''.

§2. DECOMPOSITION OVER ONE ENDOMORPHISM

Let k be a field and E a finite-dimensional vector space over k, $E \neq 0$. Let $A \in \mathrm{End}_k(E)$ be a linear map of E into itself. Let t be transcendental over k. We shall define a representation of the polynomial ring $k[t]$ in E. Namely, we have a homomorphism

$$k[t] \to k[A] \subset \mathrm{End}_k(E)$$

which is obtained by substituting A for t in polynomials. The ring $k[A]$ is the subring of $\mathrm{End}_k(E)$ generated by A, and is commutative because powers of A commute with each other. Thus if $f(t)$ is a polynomial and $v \in E$, then

$$f(t)v = f(A)v.$$

The kernel of the homomorphism $f(t) \mapsto f(A)$ is a principal ideal of $k[t]$, which is $\neq 0$ because $k[A]$ is finite dimensional over k. It is generated by a unique polynomial of degree > 0, having leading coefficient 1. This polynomial will be called the **minimal polynomial** of A over k, and will be denoted by $q_A(t)$. It is of course not necessarily irreducible.

Assume that there exists an element $v \in E$ such that $E = k[t]v = k[A]v$. This means that E is generated over k by the elements

$$v, Av, A^2v, \dots.$$

We called such a module **principal**, and if $R = k[t]$ we may write $E = Rv = (v)$.

If $q_A(t) = t^d + a_{d-1}t^{d-1} + \cdots + a_0$ then the elements

$$v, Av, \dots, A^{d-1}v$$

constitute a basis for E over k. This is proved in the same way as the analogous statement for finite field extensions. First we note that they are linearly inde pendent, because any relation of linear dependence over k would yield a poly-

nomial $g(t)$ of degree less than deg q_A and such that $g(A) = 0$. Second, they generate E because any polynomial $f(t)$ can be written $f(t) = g(t)q_A(t) + r(t)$ with deg $r <$ deg q_A. Hence $f(A) = r(A)$.

With respect to this basis, it is clear that the matrix of A is of the following type:

$$\begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & 0 & \cdots & 0 & -a_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & 0 & -a_{d-2} \\ 0 & 0 & 0 & \cdots & 1 & -a_{d-1} \end{pmatrix}.$$

If $E = (v)$ is principal, then E is isomorphic to $k[t]/(q_A(t))$ under the map $f(t) \mapsto f(A)v$. The polynomial q_A is uniquely determined by A, and does not depend on the choice of generator v for E. This is essentially obvious, because if f_1, f_2 are two polynomials with leading coefficient 1, then $k[t]/(f_1(t))$ is isomorphic to $k[t]/(f_2(t))$ if and only if $f_1 = f_2$. (Decompose each polynomial into prime powers and apply the structure theorem for modules over principal rings.)

If E is principal then we shall call the polynomial q_A above the **polynomial invariant of** E, with respect to A, or simply its **invariant**.

Theorem 2.1. *Let E be a non-zero finite-dimensional space over the field k, and let $A \in \mathrm{End}_k(E)$. Then E admits a direct sum decomposition*

$$E = E_1 \oplus \cdots \oplus E_r,$$

where each E_i is a principal $k[A]$-submodule, with invariant $q_i \neq 0$ such that

$$q_1 | q_2 | \cdots | q_r.$$

The sequence $(q_1, \ldots, q_r)$ is uniquely determined by E and A, and q_r is the minimal polynomial of A.

Proof. The first statement is simply a rephrasing in the present language for the structure theorem for modules over principal rings. Furthermore, it is clear that $q_r(A) = 0$ since $q_i | q_r$ for each i. No polynomial of lower degree than q_r can annihilate E, because in particular, such a polynomial does not annihilate E_r. Thus q_r is the minimal polynomial.

We shall call $(q_1, \ldots, q_r)$ the **invariants** of the pair (E, A). Let $E = k^{(n)}$, and let A be an $n \times n$ matrix, which we view as a linear map of E into itself. The invariants $(q_1, \ldots, q_r)$ will be called the **invariants** of A (over k).

Corollary 2.2. *Let k' be an extension field of k and let A be an $n \times n$ matrix in k. The invariants of A over k are the same as its invariants over k'.*

Proof. Let $\{v_1, \dots, v_n\}$ be a basis of $k^{(n)}$ over k. Then we may view it also as a basis of $k'^{(n)}$ over k'. (The unit vectors are in the k-space generated by $v_1, \dots, v_n$; hence $v_1, \dots, v_n$ generate the n-dimensional space $k'^{(n)}$ over k'.) Let $E = k^{(n)}$. Let L_A be the linear map of E determined by A. Let L'_A be the linear map of $k'^{(n)}$ determined by A. The matrix of L_A with respect to our given basis is the same as the matrix of L'_A. We can select the basis corresponding to the decomposition

$$E = E_1 \oplus \cdots \oplus E_r$$

determined by the invariants $q_1, \dots, q_r$. It follows that the invariants don't change when we lift the basis to one of $k'^{(n)}$.

Corollary 2.3. *Let A, B be $n \times n$ matrices over a field k and let k' be an extension field of k. Assume that there is an invertible matrix C' in k' such that $B = C'AC'^{-1}$. Then there is an invertible matrix C in k such that $B = CAC^{-1}$.*

Proof. Exercise.

The structure theorem for modules over principal rings gives us two kinds of decompositions. One is according to the invariants of the preceding theorem. The other is according to prime powers.

Let $E \neq 0$ be a finite dimensional space over the field k, and let $A : E \to E$ be in $\mathrm{End}_k(E)$. Let $q = q_A$ be its minimal polynomial. Then q has a factorization,

$$q = p_1^{e_1} \cdots p_s^{e_s} \qquad (e_i \geqq 1)$$

into prime powers (distinct). Hence E is a direct sum of submodules

$$E = E(p_1) \oplus \cdots \oplus E(p_s),$$

such that each $E(p_i)$ is annihilated by $p_i^{e_i}$. Furthermore, each such submodule can be expressed as a direct sum of submodules isomorphic to $k[t]/(p^e)$ for some irreducible polynomial p and some integer $e \geqq 1$.

Theorem 2.4. *Let $q_A(t) = (t - \alpha)^e$ for some $\alpha \in k$, $e \geqq 1$. Assume that E is isomorphic to $k[t]/(q)$. Then E has a basis over k such that the matrix of A relative to this basis is of type*

$$\begin{pmatrix} \alpha & 0 & \cdots & 0 \\ 1 & \alpha & & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & & \ddots & 0 \\ 0 & \cdots & 1 & \alpha \end{pmatrix}.$$

Proof. Since E is isomorphic to $k[t]/(q)$, there exists an element $v \in E$ such that $k[t]v = E$. This element corresponds to the unit element of $k[t]$ in the isomorphism

$$k[t]/(q) \to E.$$

We contend that the elements

$$v, (t - \alpha)v, \ldots, (t - \alpha)^{e-1}v,$$

or equivalently,

$$v, (A - \alpha)v, \ldots, (A - \alpha)^{e-1}v,$$

form a basis for E over k. They are linearly independent over k because any relation of linear dependence would yield a relation of linear dependence between

$$v, Av, \ldots, A^{e-1}v,$$

and hence would yield a polynomial $g(t)$ of degree less than $\deg q$ such that $g(A) = 0$. Since $\dim E = e$, it follows that our elements form a basis for E over k. But $(A - \alpha)^e = 0$. It is then clear from the definitions that the matrix of A with respect to this basis has the shape stated in our theorem.

Corollary 2.5. *Let k be algebraically closed, and let E be a finite-dimensional non-zero vector space over k. Let $A \in \mathrm{End}_k(E)$. Then there exists a basis of E over k such that the matrix of A with respect to this basis consists of blocks, and each block is of the type described in the theorem.*

A matrix having the form described in the preceding corollary is said to be in **Jordan canonical form**.

Remark 1. A matrix (or an endomorphism) N is said to be **nilpotent** if there exists an integer $d > 0$ such that $N^d = 0$. We see that in the decomposition of Theorem 2.4 or Corollary 2.5, the matrix M is written in the form

$$M = B + N$$

where N is nilpotent. In fact, N is a triangular matrix (i.e. it has zero coefficients on and above the diagonal), and B is a diagonal matrix, whose diagonal elements are the roots of the minimal polynomial. Such a decomposition can always be achieved whenever the field k is such that all the roots of the minimal polynomial lie in k. We observe also that the only case when the matrix N is 0 is when all the roots of the minimal polynomial have multiplicity 1. In this case, if $n = \dim E$, then the matrix M is a diagonal matrix, with n distinct elements on the diagonal.

Remark 2. The main theorem of this section can also be viewed as falling under the general pattern of decomposing a module into a direct sum as far as possible, and also giving normalized bases for vector spaces with respect to various structures, so that one can tell in a simple way the effect of an endomorphism. More formally, consider the category of pairs (E, A), consisting of a finite dimensional vector space E over a field k, and an endomorphism $A: E \to E$. By a morphism of such pairs

$$f: (E, A) \to (E', A')$$

we mean a k-homomorphism $f: E \to E'$ such that the following diagram is commutative:

$$\begin{array}{ccc} E & \xrightarrow{f} & E' \\ {\scriptstyle A}\downarrow & & \downarrow{\scriptstyle A'} \\ E & \xrightarrow[f]{} & E' \end{array}$$

It is then immediate that such pairs form a category, so we have the notion of isomorphism. One can reformulate Theorem 2.1 by stating:

Theorem 2.6. *Two pairs (E, A) and (F, B) are isomorphic if and only if they have the same invariants.*

You can prove this as Exercise 19. The Jordan basis gives a normalized form for the matrix associated with such a pair and an appropriate basis.

In the next chapter, we shall find conditions under which a normalized matrix is actually diagonal, for hermitian, symmetric, and unitary operators over the complex numbers.

As an example and application of Theorem 2.6, we prove:

Corollary 2.7. *Let k be a field and let K be a finite separable extension of degree n. Let V be a finite dimensional vector space of dimension n over k, and let $\rho, \rho' : K \to \mathrm{End}_k(V)$ be two representations of K on V; that is, embeddings of K in $\mathrm{End}_k(V)$. Then ρ, ρ' are conjugate; that is, there exists $B \in \mathrm{Aut}_k(V)$ such that*

$$\rho'(\xi) = B\rho(\xi)B^{-1} \text{ for all } \xi \in K.$$

Proof. By the primitive element theorem of field theory, there exists an element $\alpha \in K$ such that $K = k[\alpha]$. Let $p(t)$ be the irreducible polynomial of α over k. Then $(V, \rho(\alpha))$ and $(V, \rho'(\alpha))$ have the same invariant, namely $p(t)$. Hence these pairs are isomorphic by Theorem 2.6, which means that there exists $B \in \mathrm{Aut}_k(V)$ such that

$$\rho'(\alpha) = B\rho(\alpha)B^{-1}.$$

But all elements of K are linear combinations of powers of α with coefficients in k, so it follows immediately that $\rho'(\xi) = B\rho(\xi)B^{-1}$ for all $\xi \in K$, as desired.

To get a representation of K as in corollary 2.7, one may of course select a basis of K, and represent multiplication of elements of K on K by matrices with respect to this basis. In some sense, Corollary 2.7 tells us that this is the only way to get such representations. We shall return to this point of view when considering Cartan subgroups of GL_n in Chapter XVIII, §12.

§3. THE CHARACTERISTIC POLYNOMIAL

Let k be a commutative ring and E a free module of dimension n over k. We consider the polynomial ring $k[t]$, and a linear map $A : E \to E$. We have a homomorphism

$$k[t] \to k[A]$$

as before, mapping a polynomial $f(t)$ on $f(A)$, and E becomes a module over the ring $R = k[t]$. Let M be any $n \times n$ matrix in k (for instance the matrix of A relative to a basis of E). We define the **characteristic polynomial** $P_M(t)$ to be the determinant

$$\det(tI_n - M)$$

where I_n is the unit $n \times n$ matrix. It is an element of $k[t]$. Furthermore, if N is an invertible matrix in R, then

$$\det(tI_n - N^{-1}MN) = \det(N^{-1}(tI_n - M)N) = \det(tI_n - M).$$

Hence the characteristic polynomial of $N^{-1}MN$ is the same as that of M. We may therefore define the characteristic polynomial of A, and denote by P_A, the characteristic polynomial of any matrix M associated with A with respect to some basis. (If $E = 0$, we **define the characteristic polynomial to be** 1.)

If $\varphi : k \to k'$ is a homomorphism of commutative rings, and M is an $n \times n$ matrix in k, then it is clear that

$$P_{\varphi M}(t) = \varphi P_M(t)$$

where φP_M is obtained from P_M by applying φ to the coefficients of P_M.

Theorem 3.1. (Cayley-Hamilton). *We have $P_A(A) = 0$.*

Proof. Let $\{v_1, \dots, v_n\}$ be a basis of E over k. Then

$$tv_j = \sum_{i=1}^{n} a_{ij} v_i$$

where $(a_{ij}) = M$ is the matrix of A with respect to the basis. Let $\tilde{B}(t)$ be the matrix with coefficients in $k[t]$, defined in Chapter XIII, such that

$$\tilde{B}(t)B(t) = P_A(t)I_n.$$

Then

$$\tilde{B}(t)B(t)\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} P_A(t)v_1 \\ \vdots \\ P_A(t)v_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$$

because

$$B(t)\begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Hence $P_A(t)E = 0$, and therefore $P_A(A)E = 0$. This means that $P_A(A) = 0$, as was to be shown.

Assume now that k is a field. Let E be a finite-dimensional vector space over k, and let $A \in \text{End}_k(E)$. By an **eigenvector** w of A in E one means an element $w \in E$, such that there exists an element $\lambda \in k$ for which $Aw = \lambda w$. If $w \neq 0$, then λ is determined uniquely, and is called an **eigenvalue** of A. Of course, distinct eigenvectors may have the same eigenvalue.

Theorem 3.2. *The eigenvalues of A are precisely the roots of the characteristic polynomial of A.*

Proof. Let λ be an eigenvalue. Then $A - \lambda I$ is not invertible in $\text{End}_k(E)$, and hence $\det(A - \lambda I) = 0$. Hence λ is a root of P_A. The arguments are reversible, so we also get the converse.

For simplicity of notation, we often write $A - \lambda$ instead of $A - \lambda I$.

Theorem 3.3. *Let $w_1, \ldots, w_m$ be non-zero eigenvectors of A, having distinct eigenvalues. Then they are linearly independent.*

Proof. Suppose that we have

$$a_1 w_1 + \cdots + a_m w_m = 0$$

with $a_i \in k$, and let this be a shortest relation with not all $a_i = 0$ (assuming such exists). Then $a_i \neq 0$ for all i. Let $\lambda_1, \ldots, \lambda_m$ be the eigenvalues of our vectors. Apply $A - \lambda_1$ to the above relation. We get

$$a_2(\lambda_2 - \lambda_1)w_2 + \cdots + a_m(\lambda_m - \lambda_1)w_m = 0,$$

which shortens our relation, contradiction.

Corollary 3.4. *If A has n distinct eigenvalues $\lambda_1, \ldots, \lambda_n$ belonging to eigenvectors $v_1, \ldots, v_n$, and* $\dim E = n$, *then $\{v_1, \ldots, v_n\}$ is a basis for E. The matrix*

of A with respect to this basis is the diagonal matrix:

$$\begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}.$$

Warning. It is not always true that there exists a basis of E consisting of eigenvectors!

Remark. Let k be a subfield of k'. If M is a matrix in k, we can define its characteristic polynomial with respect to k, and also with respect to k'. It is clear that the characteristic polynomials thus obtained are equal. If E is a vector space over k, we shall see later how to extend it to a vector space over k'. A linear map A extends to a linear map of the extended space, and the characteristic polynomial of the linear map does not change either. Actually, if we select a basis for E over k, then $E \approx k^{(n)}$, and $k^{(n)} \subset k'^{(n)}$ in a natural way. Thus selecting a basis allows us to extend the vector space, but this seems to depend on the choice of basis. We shall give an invariant definition later.

Let $E = E_1 \oplus \cdots \oplus E_r$ be an expression of E as a direct sum of vector spaces over k. Let $A \in \text{End}_k(E)$, and assume that $AE_i \subset E_i$ for all $i = 1, \dots, r$. Then A induces a linear map on E_i. We can select a basis for E consisting of bases for $E_1, \dots, E_r$, and then the matrix for A consists of blocks. Hence we see that

$$P_A(t) = \prod_{i=1}^{r} P_{A_i}(t).$$

Thus the characteristic polynomial is multiplicative on direct sums.

Our condition above that $AE_i \subset E_i$ can also be formulated by saying that E is expressed as a $k[A]$-direct sum of $k[A]$-submodules, or also a $k[t]$-direct sum of $k[t]$-submodules. We shall apply this to the decomposition of E given in Theorem 2.1.

Theorem 3.5. *Let E be a finite-dimensional vector space over a field k, let $A \in \text{End}_k(E)$, and let $q_1, \dots, q_r$ be the invariants of (E, A). Then*

$$P_A(t) = q_1(t) \cdots q_r(t).$$

Proof. We assume that $E = k^{(n)}$ and that A is represented by a matrix M. We have seen that the invariants do not change when we extend k to a larger field, and neither does the characteristic polynomial. Hence we may assume that k is algebraically closed. In view of Theorem 2.1 we may assume that M has a

single invariant q. Write

$$q(t) = (t - \alpha_1)^{e_1} \cdots (t - \alpha_s)^{e_s}$$

with distinct $\alpha_1, \ldots, \alpha_s$. We view M as a linear map, and split out vector space further into a direct sum of submodules (over $k[t]$) having invariants

$$(t - \alpha_1)^{e_1}, \ldots, (t - \alpha_s)^{e_s}$$

respectively (this is the prime power decomposition). For each one of these submodules, we can select a basis so that the matrix of the induced linear map has the shape described in Theorem 2.4. From this it is immediately clear that the characteristic polynomial of the map having invariant $(t - \alpha)^e$ is precisely $(t - \alpha)^e$, and our theorem is proved.

Corollary 3.6. *The minimal polynomial of A and its characteristic polynomial have the same irreducible factors.*

Proof. Because q_r is the minimal polynomial, by Theorem 2.1.

We shall generalize our remark concerning the multiplicativity of the characteristic polynomial over direct sums.

Theorem 3.7. *Let k be a commutative ring, and in the following diagram,*

$$\begin{array}{ccccccccc} 0 & \longrightarrow & E' & \longrightarrow & E & \longrightarrow & E'' & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle A'} & & \downarrow{\scriptstyle A} & & \downarrow{\scriptstyle A''} & & \\ 0 & \longrightarrow & E' & \longrightarrow & E & \longrightarrow & E'' & \longrightarrow & 0 \end{array}$$

let the rows be exact sequences of free modules over k, of finite dimension, and let the vertical maps be k-linear maps making the diagram commutative. Then

$$P_A(t) = P_{A'}(t)P_{A''}(t).$$

Proof. We may assume that E' is a submodule of E. We select a basis $\{v_1, \ldots, v_m\}$ for E'. Let $\{\bar{v}_{m+1}, \ldots, \bar{v}\}$ be a basis for E'', and let $v_{m+1}, \ldots, v_n$ be elements of E mapping on $\bar{v}_{m+1}, \ldots, \bar{v}_n$ respectively. Then

$$\{v_1, \ldots, v_m, v_{m+1}, \ldots, v_n\}$$

is a basis for E (same proof as Theorem 5.2 of Chapter III), and we are in the situation discussed in §1. The matrix for A has the shape

$$\begin{pmatrix} M' & * \\ 0 & M'' \end{pmatrix}$$

where M' is the matrix for A' and M'' is the matrix for A''. Taking the characteristic polynomial with respect to this matrix obviously yields our multiplicative property.

Theorem 3.8. *Let k be a commutative ring, and E a free module of dimension n over k. Let $A \in \mathrm{End}_k(E)$. Let*

$$P_A(t) = t^n + c_{n-1}t^{n-1} + \cdots + c_0.$$

Then

$$\mathrm{tr}(A) = -c_{n-1} \quad \textit{and} \quad \det(A) = (-1)^n c_0.$$

Proof. For the determinant, we observe that $P_A(0) = c_0$. Substituting $t = 0$ in the definition of the characteristic polynomial by the determinant shows that $c_0 = (-1)^n \det(A)$.

For the trace, let M be the matrix representing A with respect to some basis, $M = (a_{ij})$. We consider the determinant $\det(tI_n - a_{ij})$. In its expansion as a sum over permutations, it will contain a diagonal term

$$(t - a_{11}) \cdots (t - a_{nn}),$$

which will give a contribution to the coefficient of t^{n-1} equal to

$$-(a_{11} + \cdots + a_{nn}).$$

No other term in this expansion will give a contribution to the coefficient of t^{n-1}, because the power of t occurring in another term will be at most t^{n-2}. This proves our assertion concerning the trace.

Corollary 3.9. *Let the notation be as in Theorem 3.7. Then*

$$\mathrm{tr}(A) = \mathrm{tr}(A') + \mathrm{tr}(A'') \quad \textit{and} \quad \det(A) = \det(A')\det(A'').$$

Proof. Clear.

We shall now interpret our results in the Euler-Grothendieck group.

Let k be a commutative ring. We consider the category whose objects are pairs (E, A), where E is a k-module, and $A \in \mathrm{End}_k(E)$. We define a morphism

$$(E', A') \to (E, A)$$

to be a k-linear map $E' \xrightarrow{f} E$ making the following diagram commutative:

$$\begin{array}{ccc} E' & \xrightarrow{f} & E \\ {\scriptstyle A'}\downarrow & & \downarrow{\scriptstyle A} \\ E' & \xrightarrow[f]{} & E \end{array}$$

Then we can define the kernel of such a morphism to be again a pair. Indeed, let E'_0 be the kernel of $f: E' \to E$. Then A' maps E'_0 into itself because

$$fA'E'_0 = AfE'_0 = 0.$$

We let A'_0 be the restriction of A' on E'_0. The pair (E'_0, A'_0) is defined to be the kernel of our morphism.

We shall denote by f again the morphism of the pair $(E', A') \to (E, A)$. We can speak of an exact sequence

$$(E', A') \to (E, A) \to (E'', A''),$$

meaning that the induced sequence

$$E' \to E \to E''$$

is exact. We also write 0 instead of (0, 0), according to our universal convention to use the symbol 0 for all things which behave like a zero element.

We observe that our pairs now behave formally like modules, and they in fact form an abelian category.

Assume that k is a field. Let $\mathfrak{A}$ consist of all pairs (E, A) where E is finite dimensional over k.

Then Theorem 3.7 asserts that the characteristic polynomial is an Euler-Poincaré map defined for each object in our category $\mathfrak{A}$, with values into the multiplicative monoid of polynomials with leading coefficient 1.

Since the values of the map are in a monoid, this generalizes slightly the notion of Chapter III, §8, when we took the values in a group. Of course when k is a field, which is the most frequent application, we can view the values of our map to be in the multiplicative group of non-zero rational functions, so our previous situation applies.

A similar remark holds now for the trace and the determinant. *If k is a field, the trace is an Euler map into the additive group of the field, and the determinant is an Euler map into the multiplicative group of the field.* We note also that all these maps (like all Euler maps) are defined on the isomorphism classes of pairs, and are defined on the Euler-Grothendieck group.

Theorem 3.10. *Let k be a commutative ring, M an $n \times n$ matrix in k, and f a polynomial in $k[t]$. Assume that $P_M(t)$ has a factorization,*

$$P_M(t) = \prod_{i=1}^{n} (t - \alpha_i)$$

into linear factors over k. Then the characteristic polynomial of $f(M)$ is given by

$$P_{f(M)}(t) = \prod_{i=1}^{n} (t - f(\alpha_i)),$$

and

$$\operatorname{tr}(f(M)) = \sum_{i=1}^{n} f(\alpha_i), \qquad \det(f(M)) = \prod_{i=1}^{n} f(\alpha_i).$$

Proof. Assume first that k is a field. Then using the canonical decomposition in terms of matrices given in Theorem 2.4, we find that our assertion is immediately obvious. When k is a ring, we use a substitution argument. It is however necessary to know that if $X = (x_{ij})$ is a matrix with algebraically independent coefficients over $\mathbf{Z}$, then $P_X(t)$ has n distinct roots $y_1, \dots, y_n$ [in an algebraic closure of $\mathbf{Q}(X)$] and that we have a homomorphism

$$\mathbf{Z}[x_{ij}, y_1, \dots, y_n] \to k$$

mapping X on M and $y_1, \dots, y_n$ on $\alpha_1, \dots, \alpha_n$. This is obvious to the reader who read the chapter on integral ring extensions, and the reader who has not can forget about this part of the theorem.

EXERCISES

1. Let T be an upper triangular square matrix over a commutative ring (i.e. all the elements below and on the diagonal are 0). Show that T is nilpotent.
2. Carry out explicitly the proof that the determinant of a matrix

$$\begin{pmatrix} M_1 & & & * & * \\ 0 & M_2 & & & \\ 0 & 0 & \ddots & & * \\ \vdots & \vdots & & \ddots & \\ 0 & 0 & \cdots & 0 & M_s \end{pmatrix}$$

where each M_i is a square matrix, is equal to the product of the determinants of the matrices $M_1, \dots, M_s$.

3. Let k be a commutative ring, and let M, M' be square $n \times n$ matrices in k. Show that the characteristic polynomials of MM' and $M'M$ are equal.
4. Show that the eigenvalues of the matrix

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

in the complex numbers are ± 1, $\pm i$.

5. Let M, M' be square matrices over a field k. Let q, q' be their respective minimal polynomials. Show that the minimal polynomial of

$$\begin{pmatrix} M & 0 \\ 0 & M' \end{pmatrix}$$

is the least common multiple of q, q'.

6. Let A be a nilpotent endomorphism of a finite dimensional vector space E over the field k. Show that $\mathrm{tr}(A) = 0$.

7. Let R be a principal entire ring. Let E be a free module over R, and let $E^\vee = \mathrm{Hom}_R(E, R)$ be its dual module. Then $E^\vee$ is free of dimension n. Let F be a submodule of E. Show that $E^\vee/F^\perp$ can be viewed as a submodule of $F^\vee$, and that its invariants are the same as the invariants of F in E.

8. Let E be a finite-dimensional vector space over a field k. Let $A \in \mathrm{Aut}_k(E)$. Show that the following conditions are equivalent:
   (a) $A = I + N$, with N nilpotent.
   (b) There exists a basis of E such that the matrix of A with respect to this basis has all its diagonal elements equal to 1 and all elements above the diagonal equal to 0.
   (c) All roots of the characteristic polynomial of A (in the algebraic closure of k) are equal to 1.

9. Let k be a field of characteristic 0, and let M be an $n \times n$ matrix in k. Show that M is nilpotent if and only if $\mathrm{tr}(M^\nu) = 0$ for $1 \leqq \nu \leqq n$.

10. Generalize Theorem 3.10 to rational functions (instead of polynomials), assuming that k is a field.

11. Let E be a finite-dimensional space over the field k. Let $\alpha \in k$. Let E_α be the subspace of E generated by all eigenvectors of a given endomorphism A of E, having α as an eigenvalue. Show that every non-zero element of E_α is an eigenvector of A having α as an eigenvalue.

12. Let E be finite dimensional over the field k. Let $A \in \mathrm{End}_k(E)$. Let v be an eigenvector for A. Let $B \in \mathrm{End}_k(E)$ be such that $AB = BA$. Show that Bv is also an eigenvector for A (if $Bv \neq 0$), with the same eigenvalue.

Diagonalizable endomorphisms

Let E be a finite-dimensional vector space over a field k, and let $S \in \mathrm{End}_k(E)$. We say that S is **diagonalizable** if there exists a basis of E consisting of eigenvectors of S. The matrix of S with respect to this basis is then a diagonal matrix.

13. (a) If S is diagonalizable, then its minimal polynomial over k is of type

$$q(t) = \prod_{i=1}^{m} (t - \lambda_i),$$

   where $\lambda_1, \ldots, \lambda_m$ are distinct elements of k.
   (b) Conversely, if the minimal polynomial of S is of the preceding type, then S is diagonalizable. [*Hint*: The space can be decomposed as a direct sum of the subspaces E_{λ_i} annihilated by $S - \lambda_i$.]

(c) If S is diagonalizable, and if F is a subspace of E such that $SF \subset F$, show that S is diagonalizable as an endomorphism of F, i.e. that F has a basis consisting of eigenvectors of S.

(d) Let S, T be endomorphisms of E, and assume that S, T commute. Assume that both S, T are diagonalizable. Show that they are simultaneously diagonalizable, i.e. there exists a basis of E consisting of eigenvectors for both S and T. [*Hint*: If λ is an eigenvalue of S, and E_λ is the subspace of E consisting of all vectors v such that $Sv = \lambda v$, then $TE_\lambda \subset E_\lambda$.]

14. Let E be a finite-dimensional vector space over an algebraically closed field k. Let $A \in \text{End}_k(E)$. Show that A can be written in a unique way as a sum

$$A = S + N$$

where S is diagonalizable, N is nilpotent, and $SN = NS$. Show that S, N can be expressed as polynomials in A. [*Hint*: Let $P_A(t) = \prod (t - \lambda_i)^{m_i}$ be the factorization of $P_A(t)$ with distinct λ_i. Let E_i be the kernel of $(A - \lambda_i)^{m_i}$. Then E is the direct sum of the E_i. Define S on E so that on E_i, $Sv = \lambda_i v$ for all $v \in E_i$. Let $N = A - S$. Show that S, N satisfy our requirements. To get S as a polynomial in A, let g be a polynomial such that $g(t) \equiv \lambda_i \bmod (t - \lambda_i)^{m_i}$ for all i, and $g(t) \equiv 0 \bmod t$. Then $S = g(A)$ and $N = A - g(A)$.]

15. After you have read the section on the tensor product of vector spaces, you can easily do the following exercise. Let E, F be finite-dimensional vector spaces over an algebraically closed field k, and let $A : E \to E$ and $B : F \to F$ be k-endomorphisms of E, F, respectively. Let

$$P_A(t) = \prod (t - \alpha_i)^{n_i} \quad \text{and} \quad P_B(t) = \prod (t - \beta_j)^{m_j}$$

be the factorizations of their respectively characteristic polynomials, into distinct linear factors. Then

$$P_{A \otimes B}(t) = \prod_{i,j} (t - \alpha_i \beta_j)^{n_i m_j}.$$

[*Hint*: Decompose E into the direct sum of subspaces E_i, where E_i is the subspace of E annihilated by some power of $A - \alpha_i$. Do the same for F, getting a decomposition into a direct sum of subspaces F_j. Then show that some power of $A \otimes B - \alpha_i \beta_j$ annihilates $E_i \otimes F_j$. Use the fact that $E \otimes F$ is the direct sum of the subspaces $E_i \otimes F_j$, and that $\dim_k(E_i \otimes F_j) = n_i m_j$.]

16. Let Γ be a free abelian group of dimension $n \geqq 1$. Let Γ' be a subgroup of dimension n also. Let $\{v_1, \ldots, v_n\}$ be a basis of Γ, and let $\{w_1, \ldots, w_n\}$ be a basis of Γ'. Write

$$w_i = \sum a_{ij} v_j.$$

Show that the index $(\Gamma : \Gamma')$ is equal to the absolute value of the determinant of the matrix (a_{ij}).

17. Prove the normal basis theorem for finite extensions of a finite field.

18. Let $A = (a_{ij})$ be a square $n \times n$ matrix over a commutative ring k. Let A_{ij} be the matrix obtained by deleting the i-th row and j-th column from A. Let $b_{ij} = (-1)^{i+j} \det(A_{ji})$, and let B be the matrix (b_{ij}). Show that $\det(B) = \det(A)^{n-1}$, by reducing the problem to the case when A is a matrix with variable coefficients over the integers. Use this same method to give an alternative proof of the Cayley-Hamilton theorem, that $P_A(A) = 0$.

19. Let (E, A) and (E', A') be pairs consisting of a finite-dimensional vector space over a field k, and a k-endomorphism. Show that these pairs are isomorphic if and only if their invariants are equal.

20. (a) How many non-conjugate elements of $GL_2(\mathbf{C})$ are there with characteristic polynomial $t^3(t+1)^2(t-1)$?
(b) How many with characteristic polynomial $t^3 - 1001t$?

21. Let V be a finite dimensional vector space over $\mathbf{Q}$ and let $A: V \to V$ be a $\mathbf{Q}$-linear map such that $A^5 = \mathrm{Id}$. Assume that if $v \in V$ is such that $Av = v$, then $v = 0$. Prove that dim V is divisible by 4.

22. Let V be a finite dimensional vector space over $\mathbf{R}$, and let $A: V \to V$ be an $\mathbf{R}$-linear map such that $A^2 = -\mathrm{Id}$. Show that dim V is even, and that V is a direct sum of 2-dimensional A-invariant subspaces.

23. Let E be a finite-dimensional vector space over an algebraically closed field k. Let A, B be k-endomorphisms of E which commute, i.e. $AB = BA$. Show that A and B have a common eigenvector. [*Hint*: Consider a subspace consisting of all vectors having a fixed element of k as eigenvalue.]

24. Let V be a finite dimensional vector space over a field k. Let A be an endomorphism of V. Let $\mathrm{Tr}(A^m)$ be the trace of A^m as an endomorphism of V. Show that the following power series in the variable t are equal:

$$\exp\left(\sum_{m=1}^{\infty} -\mathrm{Tr}(A^m)\frac{t^m}{m}\right) = \det(I - tA) \quad \text{or} \quad -\frac{d}{dt}\log\det(I - tA) = \sum_{m=1}^{\infty} \mathrm{Tr}(A^m)t^m.$$

Compare with Exercise 23 of Chapter XVIII.

25. Let V, W be finite dimensional vector spaces over k, of dimension n. Let $(v, w) \mapsto \langle v, w\rangle$ be a non-singular bilinear form on $V \times W$. Let $c \in k$, and let $A: V \to V$ and $V: W \to W$ be endomorphisms such that

$$\langle Av, Bw\rangle = c\langle v, w\rangle \text{ for all } v \in V \text{ and } w \in W.$$

Show that

$$\det(A)\det(tI - B) = (-1)^n\det(cI - tA)$$

and

$$\det(A)\det(B) = c^n.$$

For an application of Exercises 24 and 25 to a context of topology or algebraic geometry, see Hartshorne's *Algebraic Geometry*, Appendix C, §4.

26. Let $G = SL_n(\mathbf{C})$ and let K be the complex unitary group. Let A be the group of diagonal matrices with positive real components on the diagonal.
(a) Show that if $g \in \mathrm{Nor}_G(A)$ (normalizer of A in G), then $\mathbf{c}(g)$ (conjugation by g) permutes the diagonal components of A, thus giving rise to a homomorphism $\mathrm{Nor}_G(A) \to W$ to the group W of permutations of the diagonal coordinates.

By definition, the kernel of the above homomorphism is the centralizer $\mathrm{Cen}_G(A)$.
(b) Show that actually all permutations of the coordinates can be achieved by elements of K, so we get an isomorphism

$$W \approx \mathrm{Nor}_G(A)/\mathrm{Cen}_G(A) \approx \mathrm{Nor}_K(A)/\mathrm{Cen}_K(A).$$

In fact, the K on the right can be taken to be the real unitary group, because permutation matrices can be taken to have real components (0 or ± 1).

CHAPTER XV

Structure of Bilinear Forms

There are three major types of bilinear forms: hermitian (or symmetric), unitary, and alternating (skew-symmetric). In this chapter, we give structure theorems giving normalized expressions for these forms with respect to suitable bases. The chapter also follows the standard pattern of decomposing an object into a direct sum of simple objects, insofar as possible.

§1. PRELIMINARIES, ORTHOGONAL SUMS

The purpose of this chapter is to go somewhat deeper into the structure theory for our three types of forms. To do this we shall assume most of the time that our ground ring is a field, and in fact a field of characteristic $\neq 2$ in the symmetric case.

We recall our three definitions. Let E be a module over a commutative ring R. Let $g : E \times E \to R$ be a map. If g is bilinear, we call g a **symmetric** form if $g(x, y) = g(y, x)$ for all $x, y \in E$. We call g **alternating** if $g(x, x) = 0$, and hence $g(x, y) = -g(y, x)$ for all x, $y \in E$. If R has an automorphism of order 2, written $a \mapsto \bar{a}$, we say that g is a **hermitian** form if it is linear in its first variable, antilinear in its second, and

$$g(x, y) = \overline{g(y, x)}.$$

We shall write $g(x, y) = \langle x, y \rangle$ if the reference to g is clear. We also occasionally write $g(x, y) = x \cdot y$ or $g(x, x) = x^2$. We sometimes call g a **scalar product**.

If $v_1, \dots, v_m \in E$, we denote by $(v_1, \dots, v_m)$ the submodule of E generated by $v_1, \dots, v_m$.

Let g be symmetric, alternating, or hermitian. Then it is clear that the left kernel of g is equal to its right kernel, and it will simply be called the **kernel** of g.

In any one of these cases, we say that g is **non-degenerate** if its kernel is 0. Assume that E is finite dimensional over the field k. The form is non-degenerate if and only if it is non-singular, i.e., induces an isomorphism of E with its dual space (anti-dual in the case of hermitian forms).

Except for the few remarks on the anti-linearity made in the previous chapter, we don't use the results of the duality in that chapter. We need only the duality over fields, given in Chapter III. Furthermore, we don't essentially meet matrices again, except for the remarks on the pfaffian in §10.

We introduce one more notation. In the study of forms on vector spaces, we shall frequently decompose the vector space into direct sums of orthogonal subspaces. If E is a vector space with a form g as above, and F, F' are subspaces, we shall write

$$E = F \perp F'$$

to mean that E is the direct sum of F and F', and that F is orthogonal (or perpendicular) to F', in other words, $x \perp y$ (or $\langle x, y \rangle = 0$) for all $x \in F$ and $y \in F'$. We then say that E is the **orthogonal sum** of F and F'. There will be no confusion with the use of the symbol $\perp$ when we write $F \perp F'$ to mean simply that F is perpendicular to F'. The context always makes our meaning clear.

Most of this chapter is devoted to giving certain orthogonal decompositions of a vector space with one of our three types of forms, so that each factor in the sum is an easily recognizable type.

In the symmetric and hermitian case, we shall be especially concerned with direct sum decompositions into factors which are 1-dimensional. Thus if $\langle\ ,\ \rangle$ is symmetric or hermitian, we shall say that $\{v_1, \dots, v_n\}$ is an **orthogonal basis** (with respect to the form) if $\langle v_i, v_j \rangle = 0$ whenever $i \neq j$. We see that an orthogonal basis gives such a decomposition. If the form is nondegenerate, and if $\{v_1, \dots, v_n\}$ is an orthogonal basis, then we see at once that $\langle v_i, v_i \rangle \neq 0$ for all i.

Proposition 1.1. *Let E be a vector space over the field k, and let g be a form of one of the three above types. Suppose that E is expressed as an orthogonal sum,*

$$E = E_1 \perp \cdots \perp E_m.$$

Then g is non-degenerate on E if and only if it is non-degenerate on each E_i. If E_i^0 is the kernel of the restriction of g to E_i, then the kernel of g in E is the orthogonal sum

$$E^0 = E_1^0 \perp \cdots \perp E_m^0.$$

Proof. Elements v, w of E can be written uniquely

$$v = \sum_{i=1}^{m} v_i, \qquad w = \sum_{i=1}^{m} w_i$$

with $v_i, w_i \in E_i$. Then

$$v \cdot w = \sum_{i=1}^{m} v_i \cdot w_i,$$

and $v \cdot w = 0$ if $v_i \cdot w_i = 0$ for each $i = 1, \dots, m$. From this our assertion is obvious.

Observe that if $E_1, \dots, E_m$ are vector spaces over k, and $g_1, \dots, g_m$ are forms on these spaces respectively, then we can define a form $g = g_1 \oplus \cdots \oplus g_m$ on the direct sum $E = E_1 \oplus \cdots \oplus E_m$; namely if v, w are written as above, then we let

$$g(v, w) = \sum_{i=1}^{m} g_i(v_i, w_i).$$

It is then clear that, in fact, we have $E = E_1 \perp \cdots \perp E_m$. We could also write $g = g_1 \perp \cdots \perp g_m$.

Proposition 1.2. *Let E be a finite-dimensional space over the field k, and let g be a form of the preceding type on E. Assume that g is non-degenerate. Let F be a subspace of E. The form is non-degenerate on F if and only if $F + F^\perp = E$, and also if and only if it is non-degenerate on $F^\perp$.*

Proof. We have (as a trivial consequence of Chapter III, §5)

$$\dim F + \dim F^\perp = \dim E = \dim(F + F^\perp) + \dim(F \cap F^\perp).$$

Hence $F + F^\perp = E$ if and only if $\dim(F \cap F^\perp) = 0$. Our first assertion follows at once. Since $F, F^\perp$ enter symmetrically in the dimension condition, our second assertion also follows.

Instead of saying that a form is non-degenerate on E, we shall sometimes say, by abuse of language, that E is non-degenerate.

Let E be a finite-dimensional space over the field k, and let g be a form of the preceding type. Let E_0 be the kernel of the form. Then we get an induced form of the same type

$$g_0 : E/E_0 \times E/E_0 \to k,$$

because $g(x, y)$ depends only on the coset of x and the coset of y modulo E_0. Furthermore, g_0 is non-degenerate since its kernel on both sides is 0.

Let E, E' be finite-dimensional vector spaces, with forms g, g' as above, respectively. A linear map $\sigma : E \to E'$ is said to be **metric** if

$$g'(\sigma x, \sigma y) = g(x, y)$$

or in the dot notation, $\sigma x \cdot \sigma y = x \cdot y$ for all $x, y \in E$. If σ is a linear isomorphism, and is metric, then we say that σ is an **isometry**.

Let E, E_0 be as above. Then we have an induced form on the factor space E/E_0. If W is a complementary subspace of E_0, in other words, $E = E_0 \oplus W$, and if we let $\sigma : E \to E/E_0$ be the canonical map, then σ is metric, and induces an isometry of W on E/E_0. This assertion is obvious, and shows that if

$$E = E_0 \oplus W'$$

is another direct sum decomposition of E, then W' is isometric to W. We know that $W \approx E/E_0$ is nondegenerate. Hence our form determines a unique non-degenerate form, up to isometry, on complementary subspaces of the kernel.

§2. QUADRATIC MAPS

Let R be a commutative ring and let E, F be R-modules. We suppress the prefix R- as usual. We recall that a bilinear map $f : E \times E \to F$ is said to be symmetric if $f(x, y) = f(y, x)$ for all $x, y \in E$.

We say that F is **without 2-torsion** if for all $y \in F$ such that $2y = 0$ we have $y = 0$. (This holds if 2 is invertible in R.)

Let $f : E \to F$ be a mapping. We shall say that f is **quadratic** (i.e. R-quadratic) if there exists a symmetric bilinear map $g : E \times E \to F$ and a linear map $h : E \to F$ such that for all $x \in E$ we have

$$f(x) = g(x, x) + h(x).$$

Proposition 2.1. *Assume that F is without 2-torsion. Let $f : E \to F$ be quadratic, expressed as above in terms of a symmetric bilinear map and a linear map. Then g, h are uniquely determined by f. For all $x, y \in E$ we have*

$$2g(x, y) = f(x + y) - f(x) - f(y).$$

Proof. If we compute $f(x + y) - f(x) - f(y)$, then we obtain $2g(x, y)$. If g_1 is symmetric bilinear, h_1 is linear, and $f(x) = g_1(x, x) + h_1(x)$, then $2g(x, y) = 2g_1(x, y)$. Since F is assumed to be without 2-torsion, it follows that $g(x, y) = g_1(x, y)$ for all $x, y \in E$, and thus that g is uniquely determined. But then h is determined by the relation

$$h(x) = f(x) - g(x, x).$$

We call g, h the bilinear and linear maps **associated** with f.

If $f : E \to F$ is a map, we define

$$\Delta f : E \times E \to F$$

by

$$\Delta f(x, y) = f(x + y) - f(x) - f(y).$$

We say that f is **homogeneous quadratic** if it is quadratic, and if its associated linear map is 0. We shall say that F is **uniquely divisible** by 2 if for each $z \in F$ there exists a unique $u \in F$ such that $2u = z$. (Again this holds if 2 is invertible in R.)

Proposition 2.2. *Let $f: E \to F$ be a map such that Δf is bilinear. Assume that F is uniquely divisible by 2. Then the map $x \mapsto f(x) - \frac{1}{2}\Delta f(x, x)$ is $\mathbf{Z}$-linear. If f satisfies the condition $f(2x) = 4f(x)$, then f is homogeneous quadratic.*

Proof. Obvious.

By a **quadratic form** on E, one means a homogeneous quadratic map $f: E \to R$, with values in R.

In what follows, we are principally concerned with symmetric bilinear forms. The quadratic forms play a secondary role.

§3. SYMMETRIC FORMS, ORTHOGONAL BASES

Let k be a field of characteristic $\neq 2$.

Let E be a vector space over k, with the symmetric form g. We say that g is a **null** form or that E is a **null** space if $\langle x, y \rangle = 0$ for all $x, y \in E$. Since we assumed that the characteristic of k is $\neq 2$, the condition $x^2 = 0$ for all $x \in E$ implies that g is a null form. Indeed,

$$4x \cdot y = (x + y)^2 - (x - y)^2.$$

Theorem 3.1. *Let E be $\neq 0$ and finite dimensional over k. Let g be a symmetric form on E. Then there exists an orthogonal basis.*

Proof. We assume first that g is non-degenerate, and prove our assertion by induction in that case. If the dimension n is 1, then our assertion is obvious.

Assume $n > 1$. Let $v_1 \in E$ be such that $v_1^2 \neq 0$ (such an element exists since g is assumed non-degenerate). Let $F = (v_1)$ be the subspace generated by v_1. Then F is non-degenerate, and by Proposition 1.2, we have

$$E = F + F^{\perp}.$$

Furthermore, $\dim F^{\perp} = n - 1$. Let $\{v_2, \dots, v_n\}$ be an orthogonal basis of $F^{\perp}$.

Then $\{v_1, \dots, v_n\}$ are pairwise orthogonal. Furthermore, they are linearly independent, for if

$$a_1 v_1 + \cdots + a_n v_n = 0$$

with $a_i \in k$ then we take the scalar product with v_i to get $a_i v_i^2 = 0$ whence $a_i = 0$ for all i.

Remark. We have shown in fact that if g is non-degenerate, and $v \in E$ is such that $v^2 \neq 0$ then we can complete v to an orthogonal basis of E.

Suppose that the form g is degenerate. Let E_0 be its kernel. We can write E as a direct sum

$$E = E_0 \oplus W$$

for some subspace W. The restriction of g to W is non-degenerate; otherwise there would be an element of W which is in the kernel of E, and $\neq 0$. Hence if $\{v_1, \dots, v_r\}$ is a basis of E_0, and $\{w_1, \dots, w_{n-r}\}$ is an orthogonal basis of W, then

$$\{v_1, \dots, v_r, w_1, \dots, w_{n-r}\}$$

is an orthogonal basis of E, as was to be shown.

Corollary 3.2. *Let $\{v_1, \dots, v_n\}$ be an orthogonal basis of E. Assume that $v_i^2 \neq 0$ for $i \leqq r$ and $v_i^2 = 0$ for $i > r$. Then the kernel of E is equal to $(v_{r+1}, \dots, v_n)$.*

Proof. Obvious.

If $\{v_1, \dots, v_n\}$ is an orthogonal basis of E and if we write

$$X = x_1 v_1 + \cdots + x_n v_n$$

with $x_i \in k$, then

$$X^2 = a_1 x_1^2 + \cdots + a_n x_n^2$$

where $a_i = \langle v_i, v_i \rangle$. In this representation of the form, we say that it is **diagonalized**. With respect to an orthogonal basis, we see at once that the associated matrix of the form is a diagonal matrix, namely

$$\begin{pmatrix} a_1 & & & & & & \\ & a_2 & & & 0 & & \\ & & \ddots & & & & \\ & & & a_r & & & \\ & 0 & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{pmatrix}.$$

Example. Note that Exercise 33 of Chapter XIII gave an interesting example of an orthogonal decomposition involving harmonic polynomials.

§4. SYMMETRIC FORMS OVER ORDERED FIELDS

Theorem 4.1. (Sylvester) *Let k be an ordered field and let E be a finite dimensional vector space over k, with a non-degenerate symmetric form g. There exists an integer $r \geqq 0$ such that, if $\{v_1, \dots, v_n\}$ is an orthogonal basis of E, then precisely r among the n elements $v_1^2, \dots, v_n^2$ are > 0, and $n - r$ among these elements are < 0.*

Proof. Let $a_i = v_i^2$, for $i = 1, \dots, n$. After renumbering the basis elements, say $a_1, \dots, a_r > 0$ and $a_i < 0$ for $i > r$. Let $\{w_1, \dots, w_n\}$ be any orthogonal basis, and let $b_i = w_i^2$. Say $b_1, \dots, b_s > 0$ and $b_j < 0$ for $j > s$. We shall prove that $r = s$. Indeed, it will suffice to prove that

$$v_1, \dots, v_r, w_{s+1}, \dots, w_n$$

are linearly independent, for then we get $r + n - s \leqq n$, whence $r \leqq s$, and $r = s$ by symmetry. Suppose that

$$x_1 v_1 + \cdots + x_r v_r + y_{s+1} w_{s+1} + \cdots + y_n w_n = 0.$$

Then

$$x_1 v_1 + \cdots + x_r v_r = -y_{s+1} w_{s+1} - \cdots - y_n w_n.$$

Squaring both sides yields

$$a_1 x_1^2 + \cdots + a_r x_r^2 = b_{s+1} y_{s+1}^2 + \cdots + b_n y_n^2.$$

The left-hand side is $\geqq 0$, and the right-hand side is $\leqq 0$. Hence both sides are equal to 0, and it follows that $x_i = y_j = 0$, in other words that our vectors are linearly independent.

Corollary 4.2. *Assume that every positive element of k is a square. Then there exists an orthogonal basis $\{v_1, \dots, v_n\}$ of E such that $v_i^2 = 1$ for $i \leqq r$ and $v_i^2 = -1$ for $i > r$, and r is uniquely determined.*

Proof. We divide each vector in an orthogonal basis by the square root of the absolute value of its square.

A basis having the property of the corollary is called **orthonormal**. If X is an element of E having coordinates $(x_1, \dots, x_n)$ with respect to this basis, then

$$X^2 = x_1^2 + \cdots + x_r^2 - x_{r+1}^2 - \cdots - x_n^2.$$

We say that a symmetric form g is **positive definite** if $X^2 > 0$ for all $X \in E$, $X \neq 0$. This is the case if and only if $r = n$ in Theorem 4.1. We say that g is **negative definite** if $X^2 < 0$ for all $X \in E$, $X \neq 0$.

Corollary 4.3. *The vector space E admits an orthogonal decomposition $E = E^+ \perp E^-$ such that g is positive definite on E^+ and negative definite on E^-. The dimension of E^+ (or E^-) is the same in all such decompositions.*

Let us now assume that the form g is positive definite and that every positive element of k is a square.

We define the **norm** of an element $v \in E$ by

$$|v| = \sqrt{v \cdot v}.$$

Then we have $|v| > 0$ if $v \neq 0$. We also have the **Schwarz inequality**

$$|v \cdot w| \leqq |v|\,|w|$$

for all $v, w \in E$. This is proved in the usual way, expanding

$$0 \leqq (av \pm bw)^2 = (av \pm bw) \cdot (av \pm bw)$$

by bilinearity, and letting $b = |v|$ and $a = |w|$. One then gets

$$\mp 2ab\, v \cdot w \leqq 2|v|^2|w|^2.$$

If $|v|$ or $|w| = 0$ our inequality is trivial. If neither is 0 we divide by $|v|\,|w|$ to get what we want.

From the Schwarz inequality, we deduce the triangle inequality

$$|v + w| \leqq |v| + |w|.$$

We leave it to the reader as a routine exercise.

When we have a positive definite form, there is a canonical way of getting an orthonormal basis, starting with an arbitrary basis $\{v_1, \dots, v_n\}$ and proceeding inductively. Let

$$v_1' = \frac{1}{|v_1|} v_1.$$

Then v_1 has norm 1. Let

$$w_2 = v_2 - (v_2 \cdot v_1')v_1',$$

and then

$$v_2' = \frac{1}{|w_2|} w_2.$$

Inductively, we let

$$w_r = v_r - (v_r \cdot v'_1)v'_1 - \cdots - (v_r \cdot v'_{r-1})v'_{r-1}$$

and then

$$v'_r = \frac{1}{|w_r|} w_r.$$

The $\{v'_1, \ldots, v'_n\}$ is an orthonormal basis. The inductive process just described is known as the **Gram-Schmidt orthogonalization.**

§5. HERMITIAN FORMS

Let k_0 be an ordered field (a subfield of the reals, if you wish) and let $k = k_0(i)$, where $i = \sqrt{-1}$. Then k has an automorphism of order 2, whose fixed field is k_0.

Let E be a finite-dimensional vector space over k. We shall deal with a hermitian form on E, i.e. a map

$$E \times E \to k$$

written

$$(x, y) \mapsto \langle x, y \rangle$$

which is k-linear in its first variable, k-anti-linear in its second variable, and such that

$$\langle x, y \rangle = \overline{\langle y, x \rangle}$$

for all $x, y \in E$.

We observe that $\langle x, x \rangle \in k_0$ for all $x \in E$. This is essentially the reason why the proofs of statements concerning symmetric forms hold essentially without change in the hermitian case. We shall now make the list of the properties which apply to this case.

Theorem 5.1. *There exists an orthogonal basis. If the form is non-degenerate, there exists an integer r having the following property. If $\{v_1, \ldots, v_n\}$ is an orthogonal basis, then precisely r among the n elements*

$$\langle v_1, v_1 \rangle, \ldots, \langle v_n, v_n \rangle$$

are > 0 and $n - r$ among these elements are < 0.

An orthogonal basis $\{v_1, \dots, v_n\}$ such that $\langle v_i, v_i \rangle = 1$ or -1 is called an **orthonormal** basis.

Corollary 5.2. *Assume that the form is non-degenerate, and that every positive element of k_0 is a square. Then there exists an orthonormal basis.*

We say that the hermitian form is **positive definite** if $\langle x, x \rangle > 0$ for all $x \in E$. We say that it is **negative definite** if $\langle x, x \rangle < 0$ for all $x \in E$, $x \neq 0$.

Corollary 5.3. *Assume that the form is non-degenerate. Then E admits an orthogonal decomposition $E = E^+ \perp E^-$ such that the form is positive definite on E^+ and negative definite on E^-. The dimension of E^+ (or E^-) is the same in all such decompositions.*

The proofs of Theorem 5.1 and its corollaries are identical with those of the analogous results for symmetric forms, and will be left to the reader.

We have the **polarization identity**, for any k-linear map $A : E \to E$, namely

$$\langle A(x + y), (x + y) \rangle - \langle A(x - y), (x - y) \rangle = 2[\langle Ax, y \rangle + \langle Ay, x \rangle].$$

If $\langle Ax, x \rangle = 0$ for all x, we replace x by ix and get

$$\langle Ax, y \rangle + \langle Ay, x \rangle = 0,$$

$$i\langle Ax, y \rangle - i\langle Ay, x \rangle = 0.$$

From this we conclude:

If $\langle Ax, x \rangle = 0$, for all x, then $A = 0$.

This is the only statement which has no analogue in the case of symmetric forms. The presence of i in one of the above linear equations is essential to the conclusion. In practice, one uses the statement in the complex case, and one meets an analogous situation in the real case when A is symmetric. Then the statement for symmetric maps is obvious.

Assume that the hermitian form is positive definite, and that every positive element of k_0 is a square.

We have the **Schwarz inequality**, namely

$$|\langle x, y \rangle|^2 \leqq \langle x, x \rangle \langle y, y \rangle$$

whose proof comes again by expanding

$$0 \leqq \langle \alpha x + \beta y, \alpha x + \beta y \rangle$$

and setting $\alpha = \langle y, y \rangle$ and $\beta = -\langle x, y \rangle$.

We define the norm of $|x|$ to be

$$|x| = \sqrt{\langle x, x \rangle}.$$

Then we get at once the triangle inequality

$$|x + y| \leqq |x| + |y|,$$

and for $\alpha \in k$,

$$|\alpha x| = |\alpha||x|.$$

Just as in the symmetric case, given a basis, one can find an orthonormal basis by the inductive procedure of subtracting successive projections. We leave this to the reader.

§6. THE SPECTRAL THEOREM (HERMITIAN CASE)

Throughout this section, we let E be a finite dimensional space over **C**, *of dimension* $\geqq 1$, *and we endow E with a positive definite hermitian form.*

Let $A : E \to E$ be a linear map (*i.e.* **C**-linear map) of E into itself. For fixed $y \in E$, the map $x \mapsto \langle Ax, y \rangle$ is a linear functional, and hence there exists a unique element $y^* \in E$ such that

$$\langle Ax, y \rangle = \langle x, y^* \rangle$$

for all $x \in E$. We define the map $A^* : E \to E$ by $A^* y = y^*$. It is immediately clear that A^* is linear, and we shall call A^* the **adjoint** of A with respect to our hermitian form.

The following formulas are trivially verified, for any linear maps A, B of E into itself:

$$(A + B)^* = A^* + B^*, \qquad A^{**} = A,$$

$$(\alpha A)^* = \bar{\alpha} A^*, \qquad (AB)^* = B^* A^*.$$

A linear map A is called **self-adjoint** (or **hermitian**) if $A^* = A$.

Proposition 6.1. *A is hermitian if and only if $\langle Ax, x \rangle$ is real for all $x \in E$.*

Proof. Let A be hermitian. Then

$$\overline{\langle Ax, x \rangle} = \overline{\langle x, Ax \rangle} = \langle Ax, x \rangle,$$

whence $\langle Ax, x \rangle$ is real. Conversely, assume $\langle Ax, x \rangle$ is real for all x. Then

$$\langle Ax, x \rangle = \overline{\langle Ax, x \rangle} = \langle x, Ax \rangle = \langle A^* x, x \rangle,$$

and consequently $\langle (A - A^*)x, x \rangle = 0$ for all x. Hence $A = A^*$ by polarization.

Let $A : E \to E$ be a linear map. An element $\xi \in E$ is called an **eigenvector** of A if there exists $\lambda \in \mathbf{C}$ such that $A\xi = \lambda\xi$. If $\xi \neq 0$, then we say that λ is an **eigenvalue** of A, belonging to ξ.

Proposition 6.2. *Let A be hermitian. Then all eigenvalues belonging to nonzero eigenvectors of A are real. If ξ, ξ' are eigenvectors $\neq 0$ having eigenvalues λ, λ' respectively, and if $\lambda \neq \lambda'$, then $\xi \perp \xi'$.*

Proof. Let λ be an eigenvalue, belonging to the eigenvector $\xi \neq 0$. Then $\langle A\xi, \xi \rangle = \langle \xi, A\xi \rangle$, and these two numbers are equal respectively to $\lambda\langle \xi, \xi \rangle$ and $\bar{\lambda}\langle \xi, \xi \rangle$. Since $\xi \neq 0$, it follows that $\lambda = \bar{\lambda}$, i.e. that λ is real. Secondly, assume that ξ, ξ' and λ, λ' are as described above. Then

$$\langle A\xi, \xi' \rangle = \lambda\langle \xi, \xi' \rangle = \langle \xi, A\xi' \rangle = \lambda'\langle \xi, \xi' \rangle,$$

from which it follows that $\langle \xi, \xi' \rangle = 0$.

Lemma 6.3. *Let $A : E \to E$ be a linear map, and* $\dim E \geqq 1$. *Then there exists at least one non-zero eigenvector of A.*

Proof. We consider $\mathbf{C}[A]$, i.e. the ring generated by A over $\mathbf{C}$. As a vector space over $\mathbf{C}$, it is contained in the ring of endomorphisms of E, which is finite dimensional, the dimension being the same as for the ring of all $n \times n$ matrices if $n = \dim E$. Hence there exists a non-zero polynomial P with coefficients in $\mathbf{C}$ such that $P(A) = 0$. We can factor P into a product of linear factors,

$$P(X) = (X - \lambda_1) \cdots (X - \lambda_m)$$

with $\lambda_j \in \mathbf{C}$. Then $(A - \lambda_1 I) \cdots (A - \lambda_m I) = 0$. Hence not all factors $A - \lambda_j I$ can be isomorphisms, and there exists $\lambda \in \mathbf{C}$ such that $A - \lambda I$ is not an isomorphism. Hence it has an element $\xi \neq 0$ in its kernel, and we get $A\xi - \lambda\xi = 0$. This shows that ξ is a non-zero eigenvector, as desired.

Theorem 6.4. (Spectral Theorem, Hermitian Case). *Let E be a non-zero finite dimensional vector space over the complex numbers, with a positive definite hermitian form. Let $A : E \to E$ be a hermitian linear map. Then E has an orthogonal basis consisting of eigenvectors of A.*

Proof. Let ξ_1 be a non-zero eigenvector, with eigenvalue λ_1, and let E_1 be the subspace generated by ξ_1. Then A maps $E_1^{\perp}$ into itself, because

$$\langle AE_1^{\perp}, \xi_1 \rangle = \langle E_1^{\perp}, A\xi_1 \rangle = \langle E_1^{\perp}, \lambda_1\xi_1 \rangle = \lambda_1\langle E_1^{\perp}, \xi_1 \rangle = 0,$$

whence $AE_1^{\perp}$ is perpendicular to ξ_1.

Since $\xi_1 \neq 0$ we have $\langle \xi_1, \xi_1 \rangle > 0$ and hence, since our hermitian form is non-degenerate (being positive definite), we have

$$E = E_1 \oplus E_1^{\perp}.$$

The restriction of our form to $E_1^\perp$ is positive definite (if $\dim E > 1$). From Proposition 6.1, we see at once that the restriction of A to $E_1^\perp$ is hermitian. Hence we can complete the proof by induction.

Corollary 6.5. *Hypotheses being as in the theorem, there exists an orthonormal basis consisting of eigenvectors of A.*

Proof. Divide each vector in an orthogonal basis by its norm.

Corollary 6.6. *Let E be a non-zero finite dimensional vector space over the complex numbers, with a positive definite hermitian form f. Let g be another hermitian form on E. Then there exists a basis of E which is orthogonal for both f and g.*

Proof. We write $f(x, y) = \langle x, y \rangle$. Since f is non-singular, being positive definite, there exists a unique hermitian linear map A such that $g(x, y) = \langle Ax, y \rangle$ for all $x, y \in E$. We apply the theorem to A, and find a basis as in the theorem, say $\{v_1, \dots, v_n\}$. Let λ_i be the eigenvalue such that $Av_i = \lambda_i v_i$. Then

$$g(v_i, v_j) = \langle Av_i, v_j \rangle = \lambda_i \langle v_i, v_j \rangle,$$

and therefore our basis is also orthogonal for g, as was to be shown.

We recall that a linear map $U : E \to E$ is **unitary** if and only if $U^* = U^{-1}$. This condition is equivalent to the property that $\langle Ux, Uy \rangle = \langle x, y \rangle$ for all elements $x, y \in E$. In other words, U is an automorphism of the form f.

Theorem 6.7. (Spectral Theorem, Unitary Case). *Let E be a non-zero finite dimensional vector space over the complex numbers, with a positive definite hermitian form. Let $U : E \to E$ be a unitary linear map. Then E has an orthogonal basis consisting of eigenvectors of U.*

Proof. Let $\xi_1 \neq 0$ be an eigenvector of U. It is immediately verified that the subspace of E orthogonal to ξ_1 is mapped into itself by U, using the relation $U^* = U^{-1}$, because if η is perpendicular to ξ_1, then

$$\langle U\eta, \xi_1 \rangle = \langle \eta, U^*\xi_1 \rangle = \langle \eta, U^{-1}\xi_1 \rangle = \langle \eta, \lambda^{-1}\xi_1 \rangle = 0.$$

Thus we can finish the proof by induction as before.

Remark. If λ is an eigenvalue of the unitary map U, then λ has necessarily absolute value 1 (because U preserves length), whence λ can be written in the form $e^{i\theta}$ with θ real, and we may view U as a rotation.

Let $A : E \to E$ be an invertible linear map. Just as one writes a non-zero complex number $z = re^{i\theta}$ with $r > 0$, there exists a decomposition of A as a product called its polar decomposition. Let $P : E \to E$ be linear. We say that P is **semipositive** if P is hermitian and we have $\langle Px, x \rangle \geqq 0$ for all $x \in E$. If we have $\langle Px, x \rangle > 0$ for all $x \neq 0$ in E then we say that P is **positive definite**. For

example, if we let $P = A^*A$ then we see that P is positive definite, because

$$\langle A^*Ax, x\rangle = \langle Ax, Ax\rangle > 0 \text{ if } x \neq 0.$$

Proposition 6.8. *Let P be semipositive. Then P has a unique semipositive square root $B : E \to E$, i.e. a semipositive linear map such that $B^2 = P$.*

Proof. For simplicity, we assume that P is positive definite. By the spectral theorem, there exists a basis of E consisting of eigenvectors. The eigenvalues must be > 0 (immediate from the condition of positivity). The linear map defined by sending each eigenvector to its multiple by the square root of the corresponding eigenvalue satisfies the required conditions. As for uniqueness, since B commutes with P because $B^2 = P$, it follows that if $\{v_1, \ldots, v_n\}$ is a basis consisting of eigenvectors for P, then each v_i is also an eigenvector for B. (Cf. Chapter XIV, Exercises 12 and 13(d).) Since a positive number has a unique positive square root, it follows that B is uniquely determined as the unique linear map whose effect on v_i is multiplication by the square root of the corresponding eigenvalue for P.

Theorem 6.9. *Let $A : E \to E$ be an invertible linear map. Then A can be written in a unique way as a product $A = UP$, where U is unitary and P is positive definite.*

Proof. Let $P = (A^*A)^{1/2}$, and let $U = AP^{-1}$. Using the defiitions, it is immediately verified that U is unitary, so we get the existence of the decomposition. As for uniqueness, suppose $A = U_1P_1$. Let

$$U_2 = PP_1^{-1} = U^{-1}U_1.$$

Then U_2 is unitary, so $U_2^*U_2 = I$. From the fact that $P^* = P$ and $P_1^* = P_1$, we conclude that $P^2 = P_1^2$. Since P, P_1 are Hermitian positive definite, it follows as in Proposition 6.8 that $P = P_1$, thus proving the theorem.

Remark. The arguments used to prove Theorem 6.9 apply in the case of Hilbert space in analysis. *Cf.* my *Real Analysis*. However, for the uniqueness, since there may not be "eigenvalues", one has to use another technique from analysis, described in that book.

As a matter of terminology, the expression $A = UP$ in Theorem 6.9 is called the **polar decomposition** of A. Of course, it does matter in what order we write the decomposition. There is also a unique decomposition $A = P_1U_1$ with P_1 positive definite and U_1 unitary (apply Theorem 6.9 to A^{-1}, and then take inverses).

§7. THE SPECTRAL THEOREM (SYMMETRIC CASE)

Let E be a finite dimensional vector space over the real numbers, and let g be a symmetric positive definite form on E. If $A : E \to E$ is a linear map, then we know

that its transpose, relative to g, is defined by the condition

$$\langle Ax, y \rangle = \langle x, {}^tAy \rangle$$

for all $x, y \in E$. We say that A is **symmetric** if $A = {}^tA$. As before, an element $\xi \in E$ is called an eigenvector of A if there exists $\lambda \in R$ such that $A\xi = \lambda\xi$, and λ is called an eigenvalue if $\xi \neq 0$.

Theorem 7.1. **(Spectral Theorem, Symmetric Case).** *Let $E \neq 0$. Let $A : E \to E$ be a symmetric linear map. Then E has an orthogonal basis consisting of eigenvectors of A.*

Proof. If we select an orthogonal basis for the positive definite form, then the matrix of A with respect to this basis is a real symmetric matrix, and we are reduced to considering the case when $E = \mathbf{R}^n$. Let M be the matrix representing A. We may view M as operating on $\mathbf{C}^n$, and then M represents a hermitian linear map. Let $z \neq 0$ be a complex eigenvector for M, and write

$$z = x + iy,$$

with $x, y \in \mathbf{R}^n$. By Proposition 6.2, we know that an eigenvalue λ for M, belonging to z, is real, and we have $Mz = \lambda z$. Hence $Mx = \lambda x$ and $My = \lambda y$. But we must have $x \neq 0$ or $y \neq 0$. Thus we have found a nonzero eigenvector for M, namely, A, in E. We can now proceed as before. The orthogonal complement of this eigenvector in E has dimension $(n - 1)$, and is mapped into itself by A. We can therefore finish the proof by induction.

Remarks. The spectral theorems are valid over a real closed field; our proofs don't need any change. Furthermore, the proofs are reasonably close to those which would be given in analysis for Hilbert spaces, and compact operators. The existence of eigenvalues and eigenvectors must however be proved differently, for instance using the Gelfand-Mazur theorem which we have actually proved in Chapter XII, or using a variational principle (i.e. finding a maximum or minimum for the quadratic function depending on the operator).

Corollary 7.2. *Hypotheses being as in the theorem, there exists an orthonormal basis consisting of eigenvectors of A.*

Proof. Divide each vector in an orthogonal basis by its norm.

Corollary 7.3. *Let E be a non-zero finite dimensional vector space over the reals, with a positive definite symmetric form f. Let g be another symmetric form on E. Then there exists a basis of E which is orthogonal for both f and g.*

Proof. We write $f(x, y) = \langle x, y \rangle$. Since f is non-singular, being positive definite, there exists a unique symmetric linear map A such that

$$g(x, y) = \langle Ax, y \rangle$$

for all $x, y \in E$. We apply the theorem to A, and find a basis as in the theorem. It is clearly an orthogonal basis for g (cf. the same proof in the hermitian case).

The analogues of Proposition 6.8 and the polar decomposition also hold in the present case, with the same proofs. See Exercise 9.

§8. ALTERNATING FORMS

Let E be a vector space over the field k, on which we now make no restriction. We let f be an alternating form on E, i.e. a bilinear map $f: E \times E \to k$ such that $f(x, x) = x^2 = 0$ for all $x \in E$. Then

$$x \cdot y = -y \cdot x$$

for all $x, y \in E$, as one sees by substituting $(x + y)$ for x in $x^2 = 0$.

We define a **hyperbolic plane** (for the alternating form) to be a 2-dimensional space which is non-degenerate. We get automatically an element w such that $w^2 = 0$, $w \neq 0$. If P is a hyperbolic plane, and $w \in P$, $w \neq 0$, then there exists an element $y \neq 0$ in P such that $w \cdot y \neq 0$. After dividing y by some constant, we may assume that $w \cdot y = 1$. Then $y \cdot w = -1$. Hence the matrix of the form with respect to the basis $\{w, y\}$ is

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

The pair w, y is called a **hyperbolic pair** as before. Given a 2-dimensional vector space over k with a bilinear form, and a pair of elements $\{w, y\}$ satisfying the relations

$$w^2 = y^2 = 0, \qquad y \cdot w = -1, \qquad w \cdot y = 1,$$

then we see that the form is alternating, and that (w, y) is a hyperbolic plane for the form.

Given an alternating form f on E, we say that E (or f) is **hyperbolic** if E is an orthogonal sum of hyperbolic planes. We say that E (or f) is **null** if $x \cdot y = 0$ for all $x, y \in E$.

Theorem 8.1. *Let f be an alternating form on the finite dimensional vector space E over k. Then E is an orthogonal sum of its kernel and a hyperbolic subspace. If E is non-degenerate, then E is a hyperbolic space, and its dimension is even.*

Proof. A complementary subspace to the kernel is non-degenerate, and hence we may assume that E is non-degenerate. Let $w \in E$, $w \neq 0$. There exists $y \in E$ such that $w \cdot y \neq 0$ and $y \neq 0$. Then (w, y) is non-degenerate, hence is a hyperbolic plane P. We have $E = P \oplus P^{\perp}$ and $P^{\perp}$ is non-degenerate. We

complete the proof by induction.

Corollary 8.2. *All alternating non-degenerate forms of a given dimension over a field k are isometric.*

We see from Theorem 8.1 that there exists a basis of E such that relative to this basis, the matrix of the alternating form is

$$\begin{pmatrix} 0 & 1 & & & & & & & & \\ -1 & 0 & & & & & & & & \\ & & 0 & 1 & & & & & & \\ & & -1 & 0 & & & & & & \\ & & & & \ddots & & & & & \\ & & & & & 0 & 1 & & & \\ & & & & & -1 & 0 & & & \\ & & & & & & & 0 & & \\ & & & & & & & & \ddots & \\ & & & & & & & & & 0 \end{pmatrix}.$$

For convenience of writing, we reorder the basis elements of our orthogonal sum of hyperbolic planes in such a way that the matrix of the form is

$$\begin{pmatrix} 0 & I_r & 0 \\ -I_r & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

where I_r is the unit $r \times r$ matrix. The matrix

$$\begin{pmatrix} 0 & I_r \\ -I_r & 0 \end{pmatrix}$$

is called the **standard alternating** matrix.

Corollary 8.3. *Let E be a finite dimensional vector space over k, with a non-degenerate symmetric form denoted by $\langle\ ,\ \rangle$. Let Ω be a non-degenerate alternating form on E. Then there exists a direct sum decomposition $E = E_1 \oplus E_2$ and a symmetric automorphism A of E (with respect to $\langle\ ,\ \rangle$) having the following property. If $x, y \in E$ are written*

$$x = (x_1, x_2) \quad \textit{with} \quad x_1 \in E_1 \quad \textit{and} \quad x_2 \in E_2,$$

$$y = (y_1, y_2) \quad \textit{with} \quad y_1 \in E_1 \quad \textit{and} \quad y_2 \in E_2,$$

then

$$\Omega(x, y) = \langle Ax_1, y_2 \rangle - \langle Ax_2, y_1 \rangle.$$

Proof. Take a basis of E such that the matrix of Ω with respect to this basis is the standard alternating matrix. Let f be the symmetric non-degenerate form on E given by the dot product with respect to this basis. Then we obtain a direct sum decomposition of E into subspaces E_1, E_2 (corresponding to the first n, resp. the last n coordinates), such that

$$\Omega(x, y) = f(x_1, y_2) - f(x_2, y_1).$$

Since $\langle\ ,\ \rangle$ is assumed non-degenerate, we can find an automorphism A having the desired effect, and A is symmetric because f is symmetric.

§9. THE PFAFFIAN

An alternating matrix is a matrix G such that ${}^tG = -G$ and the diagonal elements are equal to 0. As we saw in Chapter XIII, §6, it is the matrix of an alternating form. We let G be an $n \times n$ matrix, and assume n is *even*. (For odd n, cf. exercises.)

We start over a field of characteristic 0. By Corollary 8.2, there exists a non-singular matrix C such that ${}^tCGC$ is the matrix

$$\begin{pmatrix} 0 & I_r & 0 \\ -I_r & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

and hence

$$\det(C)^2 \det(G) = 1 \quad \text{or} \quad 0$$

according as the kernel of the alternating form is trivial or non-trivial. Thus in any case, we see that $\det(G)$ is a square in the field.

Now we move over to the integers $\mathbf{Z}$. Let t_{ij} ($1 \leqq i < j \leqq n$) be $n(n-1)/2$ algebraically independent elements over $\mathbf{Q}$, let $t_{ii} = 0$ for $i = 1, \ldots, n$, and let $t_{ij} = -t_{ji}$ for $i > j$. Then the matrix $T = (t_{ij})$ is alternating, and hence $\det(T)$ is a square in the field $\mathbf{Q}(t)$ obtained from $\mathbf{Q}$ by adjoining all the variables t_{ij}. However, $\det(T)$ is a polynomial in $\mathbf{Z}[t]$, and since we have unique factorization in $\mathbf{Z}[t]$, it follows that $\det(T)$ is the square of a polynomial in $\mathbf{Z}[t]$. We can write

$$\det(T) = P(t)^2.$$

The polynomial P is uniquely determined up to a factor of ± 1. If we substitute

values for the t_{ij} so that the matrix T specializes to

$$\begin{pmatrix} 0 & I_{n/2} \\ -I_{n/2} & 0 \end{pmatrix},$$

then we see that there exists a unique polynomial P with integer coefficients taking the value 1 for this specialized set of values of (t). We call P the **generic Pfaffian** of size n, and write it Pf.

Let R be a commutative ring. We have a homomorphism

$$\mathbf{Z}[t] \to R[t]$$

induced by the unique homomorphism of $\mathbf{Z}$ into R. The image of the generic Pfaffian of size n in $R[t]$ is a polynomial with coefficients in R, which we still denote by Pf. If G is an alternating matrix with coefficients in R, then we write $\text{Pf}(G)$ for the value of $\text{Pf}(t)$ when we substitute g_{ij} for t_{ij} in Pf. Since the determinant commutes with homomorphisms, we have:

Theorem 9.1. *Let R be a commutative ring. Let $(g_{ij}) = G$ be an alternating matrix with $g_{ij} \in R$. Then*

$$\det(G) = (\text{Pf}(G))^2.$$

Furthermore, if C is an $n \times n$ matrix in R, then

$$\text{Pf}(CG^tC) = \det(C)\,\text{Pf}(G).$$

Proof. The first statement has been proved above. The second statement will follow if we can prove it over $\mathbf{Z}$. Let u_{ij} $(i, j = 1, \ldots, n)$ be algebraically independent over $\mathbf{Q}$, and such that u_{ij}, t_{ij} are algebraically independent over $\mathbf{Q}$. Let U be the matrix (u_{ij}). Then

$$\text{Pf}(UT^tU) = \pm\det(U)\,\text{Pf}(T),$$

as follows immediately from taking the square of both sides. Substitute values for U and T such that U becomes the unit matrix and T becomes the standard alternating matrix. We conclude that we must have a $+$ sign on the right-hand side. Our assertion now follows as usual for any substitution of U to a matrix in R, and any substitution of T to an alternating matrix in R, as was to be shown.

§10. WITT'S THEOREM

We go back to symmetric forms and we let k be a field of characteristic $\neq 2$.

Let E be a vector space over k, with a symmetric form. We say that E is a **hyperbolic plane** if the form is non-degenerate, if E has dimension 2, and if there exists an element $w \neq 0$ in E such that $w^2 = 0$. We say that E is a **hyperbolic space** if it is an orthogonal sum of hyperbolic planes. We also say that the form on E is hyperbolic.

Suppose that E is a hyperbolic plane, with an element $w \neq 0$ such that $w^2 = 0$. Let $u \in E$ be such that $E = (w, u)$. Then $u \cdot w \neq 0$; otherwise w would be a non-zero element in the kernel. Let $b \in k$ be such that $w \cdot bu = bw \cdot u = 1$.

Then select $a \in k$ such that

$$(aw + bu)^2 = 2abw \cdot u + b^2u^2 = 0.$$

(This can be done since we deal with a linear equation in a.) Put $v = aw + bu$. Then we have found a basis for E, namely $E = (w, v)$ such that

$$w^2 = v^2 = 0 \quad \text{and} \quad w \cdot v = 1.$$

Relative to this basis, the matrix of our form is therefore

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

We observe that, conversely, a space E having a basis $\{w, v\}$ satisfying $w^2 = v^2 = 0$ and $w \cdot v = 1$ is non-degenerate, and thus is a hyperbolic plane. A basis $\{w, v\}$ satisfying these relations will be called a **hyperbolic pair**.

An orthogonal sum of non-degenerate spaces is non-degenerate and hence a hyperbolic space is non-degenerate. We note that a hyperbolic space always has even dimension.

Lemma 10.1. *Let E be a finite dimensional vector space over k, with a non-degenerate symmetric form g. Let F be a subspace, F_0 the kernel of F, and suppose we have an orthogonal decomposition*

$$F = F_0 \perp U.$$

Let $\{w_1, \dots, w_s\}$ be a basis of F_0. Then there exist elements $v_1, \dots, v_s$ in E perpendicular to U, such that each pair $\{w_i, v_i\}$ is a hyperbolic pair generating a hyperbolic plane P_i, and such that we have an orthogonal decomposition

$$U \perp P_1 \perp \cdots \perp P_s.$$

Proof. Let

$$U_1 = (w_2, \dots, w_s) \oplus U.$$

Then U_1 is contained in $F_0 \oplus U$ properly, and consequently $(F_0 \oplus U)^\perp$ is

contained in $U_1^\perp$ properly. Hence there exists an element $u_1 \in U_1^\perp$ but

$$u_1 \notin (F_0 \oplus U)^\perp.$$

We have $w_1 \cdot u_1 \neq 0$, and hence (w_1, u_1) is a hyperbolic plane P_1. We have seen previously that we can find $v_1 \in P_1$ such that $\{w_1, v_1\}$ is a hyperbolic pair. Furthermore, we obtain an orthogonal sum decomposition

$$F_1 = (w_2, \ldots, w_s) \perp P_1 \perp U.$$

Then it is clear that $(w_2, \ldots, w_s)$ is the kernel of F_1, and we can complete the proof by induction.

Theorem 10.2 *Let E be a finite dimensional vector space over k, and let g be a non-degenerate symmetric form on E. Let F, F' be subspaces of E, and let $\sigma: F \to F'$ be an isometry. Then σ can be extended to an isometry of E onto itself.*

Proof. We shall first reduce the proof to the case when F is non-degenerate.

We can write $F = F_0 \perp U$ as in the lemma of the preceding section, and then $\sigma F = F' = \sigma F_0 \perp \sigma U$. Furthermore, $\sigma F_0 = F_0'$ is the kernel of F'. Now we can enlarge both F and F' as in the lemma to orthogonal sums

$$U \perp P_1 \perp \cdots \perp P_s \quad \text{and} \quad \sigma U \perp P_1' \perp \cdots \perp P_s'$$

corresponding to a choice of basis in F_0 and its corresponding image in F_0'. Thus we can extend σ to an isometry of these extended spaces, which are non-degenerate. This gives us the desired reduction.

We assume that F, F' are non-degenerate, and proceed stepwise.

Suppose first that $F' = F$, i.e. that σ is an isometry of F onto itself. We can extend σ to E simply by leaving every element of $F^\perp$ fixed.

Next, assume that $\dim F = \dim F' = 1$ and that $F \neq F'$. Say $F = (v)$ and $F' = (v')$. Then $v^2 = v'^2$. Furthermore, (v, v') has dimension 2.

If (v, v') is non-degenerate, it has an isometry extending σ, which maps v on v' and v' on v. We can apply the preceding step to conclude the proof.

If (v, v') is degenerate, its kernel has dimension 1. Let w be a basis for this kernel. There exist $a, b \in k$ such that $v' = av + bw$. Then $v'^2 = a^2 v^2$ and hence $a = \pm 1$. Replacing v' by $-v'$ if necessary, we may assume $a = 1$. Replacing w by bw, we may assume $v' = v + w$. Let $z = v + v'$. We apply Lemma 10.1 to the space

$$(w, z) = (w) \perp (z).$$

We can find an element $y \in E$ such that

$$y \cdot z = 0, \qquad y^2 = 0, \quad \text{and} \quad w \cdot y = 1.$$

The space $(z, w, y) = (z) \perp (w, y)$ is non-degenerate, being an orthogonal sum of (z) and the hyperbolic plane (w, y). It has an isometry such that

$$z \leftrightarrow z, \qquad w \leftrightarrow -w, \qquad y \leftrightarrow -y.$$

But $v = \frac{1}{2}(z - w)$ is mapped on $v' = \frac{1}{2}(z + w)$ by this isometry. We have settled the present case.

We finish the proof by induction. By the existence of an orthogonal basis (Theorem 3.1), every subspace F of dimension > 1 has an orthogonal decomposition into a sum of subspaces of smaller dimension. Let $F = F_1 \perp F_2$ with $\dim F_1$ and $\dim F_2 \geqq 1$. Then

$$\sigma F = \sigma F_1 \perp \sigma F_2.$$

Let $\sigma_1 = \sigma | F_1$ be the restriction of σ to F_1. By induction, we can extend σ_1 to an isometry

$$\bar{\sigma}_1 : E \to E.$$

Then $\bar{\sigma}_1(F_1^\perp) = (\sigma_1 F_1)^\perp$. Since σF_2 is perpendicular to $\sigma F_1 = \sigma_1 F_1$, it follows that σF_2 is contained in $\bar{\sigma}_1(\mathrm{F}_1^\perp)$. Let $\sigma_2 = \sigma | F_2$. Then the isometry

$$\sigma_2 : F_2 \to \sigma_2 F_2 = \sigma F_2$$

extends by induction to an isometry

$$\bar{\sigma}_2 : F_1^\perp \to \bar{\sigma}_1(F_1^\perp).$$

The pair $(\sigma_1, \bar{\sigma}_2)$ gives us an isometry of $F_1 \perp F_1^\perp = E$ onto itself, as desired.

Corollary 10.3. *Let E, E' be finite dimensional vector spaces with non-degenerate symmetric forms, and assume that they are isometric. Let F, F' be subspaces, and let $\sigma : F \to F'$ be an isometry. Then σ can be extended to an isometry of E onto E'.*

Proof. Clear.

Let E be a space with a symmetric form g, and let F be a null subspace. Then by Lemma 10.1, we can embed F in a hyperbolic subspace H whose dimension is $2 \dim F$.

As applications of Theorem 10.2, we get several corollaries.

Corollary 10.4. *Let E be a finite dimensional vector space with a non-degenerate symmetric form. Let W be a maximal null subspace, and let W' be some null subspace. Then* $\dim W' \leqq \dim W$, *and W' is contained in some maximal null subspace, whose dimension is the same as* $\dim W$.

Proof. That W' is contained in a maximal null subspace follows by Zorn's lemma. Suppose $\dim W' \geqq \dim W$. We have an isometry of W onto a subspace of W' which we can extend to an isometry of E onto itself. Then $\sigma^{-1}(W')$ is a null subspace containing W, hence is equal to W, whence $\dim W = \dim W'$. Our assertions follow by symmetry.

Let E be a vector space with a non-degenerate symmetric form. Let W be a null subspace. By Lemma 10.1 we can embed W in a hyperbolic subspace H of E such that W is the maximal null subspace of H, and H is non-degenerate. Any such H will be called a **hyperbolic enlargement** of W.

Corollary 10.5. *Let E be a finite dimensional vector space with a non-degenerate symmetric form. Let W and W' be maximal null subspaces. Let H, H' be hyperbolic enlargements of W, W' respectively. Then H, H' are isometric and so are $H^{\perp}$ and $H'^{\perp}$.*

Proof. We have obviously an isometry of H on H', which can be extended to an isometry of E onto itself. This isometry maps $H^{\perp}$ on $H'^{\perp}$, as desired.

Corollary 10.6. *Let g_1, g_2, h be symmetric forms on finite dimensional vector spaces over the field of k. If $g_1 \oplus h$ is isometric to $g_2 \oplus h$, and if g_1, g_2 are non-degenerate, then g_1 is isometric to g_2.*

Proof. Let g_1 be a form on E_1 and g_2 a form on E_2. Let h be a form on F. Then we have an isometry between $F \oplus E_1$ and $F \oplus E_2$. Extend the identity $\mathrm{id}: F \to F$ to an isometry σ of $F \oplus E_1$ to $F \oplus E_2$ by Corollary 10.3. Since E_1 and E_2 are the respective orthogonal complements of F in their two spaces, we must have $\sigma(E_1) = E_2$, which proves what we wanted.

If g is a symmetric form on E, we shall say that g is **definite** if $g(x, x) \neq 0$ for any $x \in E$, $x \neq 0$ (i.e. $x^2 \neq 0$ if $x \neq 0$).

Corollary 10.7. *Let g be a symmetric form on E. Then g has a decomposition as an orthogonal sum*

$$g = g_0 \oplus g_{\mathrm{hyp}} \oplus g_{\mathrm{def}}$$

where g_0 is a null form, g_{hyp} is hyperbolic, and g_{def} is definite. The form $g_{\mathrm{hyp}} \oplus g_{\mathrm{def}}$ is non-degenerate. The forms g_0, g_{hyp}, and g_{def} are uniquely determined up to isometries.

Proof. The decomposition $g = g_0 \oplus g_1$ where g_0 is a null form and g_1 is non-degenerate is unique up to an isometry, since g_0 corresponds to the kernel of g.

We may therefore assume that g is non-degenerate. If

$$g = g_h \oplus g_d$$

where g_h is hyperbolic and g_d is definite, then g_h corresponds to the hyperbolic enlargement of a maximal null subspace, and by Corollary 10.5 it follows that g_h is uniquely determined. Hence g_d is uniquely determined as the orthogonal complement of g_h. (By uniquely determined, we mean of course up to an isometry.)

We shall abbreviate g_{hyp} by g_h and g_{def} by g_d.

§11. THE WITT GROUP

Let g, φ by symmetric forms on finite dimensional vector spaces over k. We shall say that they are **equivalent** if g_d is isometric to φ_d. The reader will verify at once that this is an equivalence relation. Furthermore the (orthogonal) sum of two null forms is a null form, and the sum of two hyperbolic forms is hyperbolic. However, the sum of two definite forms need not be definite. We write our equivalence $g \sim \varphi$. Equivalence is preserved under orthogonal sums, and hence equivalence classes of symmetric forms constitute a monoid.

Theorem 11.1. *The monoid of equivalence classes of symmetric forms (over the field k) is a group.*

Proof. We have to show that every element has an additive inverse. Let g be a symmetric form, which we may assume definite. We let $-g$ be the form such that $(-g)(x, y) = -g(x, y)$. We contend that $g \oplus -g$ is equivalent to 0. Let E be the space on which g is defined. Then $g \oplus -g$ is defined on $E \oplus E$. Let W be the subspace consisting of all pairs (x, x) with $x \in E$. Then W is a null space for $g \oplus -g$. Since $\dim(E \oplus E) = 2 \dim W$, it follows that W is a maximal null space, and that $g \oplus -g$ is hyperbolic, as was to be shown.

The group of Theorem 11.1 will be called the **Witt group** of k, and will be denoted by $W(k)$. It is of importance in the study of representations of elements of k by the quadratic form f arising from g [i.e. $f(x) = g(x, x)$], for instance when one wants to classify the definite forms f.

We shall now define another group, which is of importance in more functorial studies of symmetric forms, for instance in studying the quadratic forms arising from manifolds in topology.

We observe that isometry classes of non-degenerate symmetric forms (over k) constitute a monoid $M(k)$, the law of composition being the orthogonal sum. Furthermore, the cancellation law holds (Corollary 10.6). We let

$$\text{cl} : M(k) \to WG(k)$$

be the canonical map of $M(k)$ into the Grothendieck group of this monoid, which we shall call the **Witt-Grothendieck** group over k. As we know, the cancellation law implies that cl is injective.

If g is a symmetric non-degenerate form over k, we define its dimension $\dim g$ to be the dimension of the space E on which it is defined. Then it is clear that

$$\dim(g \oplus g') = \dim g + \dim g'.$$

Hence dim factors through a homomorphism

$$\dim : WG(k) \to \mathbf{Z}.$$

This homomorphism splits since we have a non-degenerate symmetric form of dimension 1.

Let $WG_0(k)$ be the kernel of our homomorphism dim. If g is a symmetric non-degenerate form we can define its determinant $\det(g)$ to be the determinant of a matrix G representing g relative to a basis, modulo squares. This is well defined as an element of k^*/k^{*2}. We define det of the 0-form to be 1. Then det is a homomorphism

$$\det : M(k) \to k^*/k^{*2},$$

and can therefore be factored through a homomorphism, again denoted by det, of the Witt-Grothendieck group, $\det : WG(k) \to k^*/k^{*2}$.

Other properties of the Witt-Grothendieck group will be given in the exercises.

EXERCISES

1. (a) Let E be a finite dimensional space over the complex numbers, and let

$$h : E \times E \to \mathbf{C}$$

be a hermitian form. Write

$$h(x, y) = g(x, y) + if(x, y)$$

where g, f are real valued. Show that g, f are $\mathbf{R}$-bilinear, g is symmetric, f is alternating.

(b) Let E be finite dimensional over $\mathbf{C}$. Let $g : E \times E \to \mathbf{C}$ be $\mathbf{R}$-bilinear. Assume that for all $x \in E$, the map $y \mapsto g(x, y)$ is $\mathbf{C}$-linear, and that the R-bilinear form

$$f(x, y) = g(x, y) - g(y, x)$$

is real-valued on $E \times E$. Show that there exists a hermitian form h on E and a symmetric $\mathbf{C}$-bilinear form ψ on E such that $2ig = h + \psi$. Show that h and ψ are uniquely determined.

2. Prove the real case of the unitary spectral theorem: If E is a non-zero finite dimensional space over $\mathbf{R}$, with a positive definite symmetric form, and $U : E \to E$ is a unitary linear map, then E has an orthogonal decomposition into subspaces of dimension 1 or 2, invariant under U. If $\dim E = 2$, then the matrix of U with respect to any orthonormal basis is of the form

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

depending on whether $\det(U) = 1$ or -1. Thus U is a rotation, or a rotation followed by a reflection.

3. Let E be a finite-dimensional, non-zero vector space over the reals, with a positive definite scalar product. Let $T : E \to E$ be a unitary automorphism of E. Show that E is an orthogonal sum of subspaces

$$E = E_1 \perp \cdots \perp E_m$$

such that each E_i is T-invariant, and has dimension 1 or 2. If E has dimension 2, show that one can find a basis such that the matrix associated with T with respect to this basis is

$$\begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} -\cos\theta & \sin\theta \\ \sin\theta & \cos\theta \end{pmatrix},$$

according as $\det T = 1$ or $\det T = -1$.

4. Let E be a finite dimensional non-zero vector space over $\mathbf{C}$, with a positive definite hermitian product. Let $A, B : E \to E$ be a hermitian endomorphism. Assume that $AB = BA$. Prove that there exists a basis of E consisting of common eigenvectors for A and B.

5. Let E be a finite-dimensional space over the complex, with a positive definite hermitian form. Let S be a set of ($\mathbf{C}$-linear) endomorphisms of E having no invariant subspace except 0 and E. (This means that if F is a subspace of E and $BF \subset F$ for all $B \in S$, then $F = 0$ or $F = E$.) Let A be a hermitian map of E into itself such that $AB = BA$ for all $B \in S$. Show that $A = \lambda I$ for some real number λ. [*Hint*: Show that there exists exactly one eigenvalue of A. If there were two eigenvalues, say $\lambda_1 \neq \lambda_2$, one could find two polynomials f and g with real coefficients such that $f(A) \neq 0$, $g(A) \neq 0$ but $f(A)g(A) = 0$. Let F be the kernel of $g(A)$ and get a contradiction.]

6. Let E be as in Exercise 5. Let T be a $\mathbf{C}$-linear map of E into itself. Let

$$A = \tfrac{1}{2}(T + T^*).$$

Show that A is hermitian. Show that T can be written in the form $A + iB$ where A, B are hermitian, and are uniquely determined.

7. Let S be a commutative set of $\mathbf{C}$-linear endomorphisms of E having no invariant subspace unequal to 0 or E. Assume in addition that if $B \in S$, then $B^* \in S$. Show that each

element of S is of type αI for some complex number α. [*Hint*: Let $B_0 \in S$. Let

$$A = \frac{1}{2}(B_0 + B_0^*).$$

Show that $A = \lambda I$ for some real λ.]

8. An endomorphism B of E is said to be **normal** if B commutes with B^*. State and prove a spectral theorem for normal endomorphisms.

Symmetric endomorphisms

For Exercises 9, 10 and 11 we let E be a non-zero finite dimensional vector space over $\mathbf{R}$, *with a symmetric positive definite scalar product g, which gives rise to a norm $|\ |$ on E.*

Let $A : E \to E$ be a symmetric endomorphism of E with respect to g. Define $A \geqq 0$ to mean $\langle Ax, x \rangle \geqq 0$ for all $x \in E$.

9. (a) Show that $A \geqq 0$ if and only if all eigenvalues of A belonging to non-zero eigenvectors are $\geqq 0$. Both in the hermitian case and the symmetric case, one says that A is **semipositive** if $A \geqq 0$, and **positive definite** if $\langle Ax, x \rangle > 0$ for all $x \neq 0$.
   (b) Show that an automorphism A of E can be written in a unique way as a product $A = UP$ where U is real unitary (that is, ${}^tUU = I$), and P is symmetric positive definite. For two hermitian or symmetric endomorphisms A, B, define $A \geqq B$ to mean $A - B \geqq 0$, and similarly for $A > B$. Suppose $A > 0$. Show that there are two real numbers $\alpha > 0$ and $\beta > 0$ such that $\alpha I \leqq A \leqq \beta I$.

10. If A is an endomorphism of E, define its norm $|A|$ to be the greatest lower bound of all numbers C such that $|Ax| \leqq C|x|$ for all $x \in E$.
    (a) Show that this norm satisfies the triangle inequality.
    (b) Show that the series

    $$\exp(A) = I + A + \frac{A^2}{2!} + \cdots$$

    converges, and if A commutes with B, then $\exp(A + B) = \exp(A)\exp(B)$. If A is sufficiently close to I, show that the series

    $$\log(A) = \frac{(A - I)}{1} - \frac{(A - I)^2}{2} + \cdots$$

    converges, and if A commutes with B, then

    $$\log(AB) = \log A + \log B.$$

    (c) Using the spectral theorem, show how to define $\log P$ for arbitrary positive definite endomorphisms P.

11. Again, let E be non-zero finite dimensional over $\mathbf{R}$, and with a positive definite symmetric form. Let $A : E \to E$ be a linear map. Prove:
    (a) If A is symmetric (resp. alternating), then $\exp(A)$ is symmetric positive definite (resp. real unitary).
    (b) If A is a linear automorphism of E sufficiently close to I, and is symmetric

positive definite (resp. real unitary), then $\log A$ is symmetric (resp. alternating).

(c) More generally, if A is positive definite, then $\log A$ is symmetric.

12. Let R be a commutative ring, let E, F be R-modules, and let $f: E \to F$ be a mapping. Assume that multiplication by 2 in F is an invertible map. Show that f is homogeneous quadratic if and only if f satisfies the **parallelogram law**:

$$f(x + y) + f(x - y) = 2f(x) + 2f(y)$$

for all $x, y \in E$.

13. (Tate) Let E, F be complete normed vector spaces over the real numbers. Let $f: E \to F$ be a map having the following property. There exists a number $C > 0$ such that for all $x, y \in E$ we have

$$|f(x + y) - f(x) - f(y)| \leqq C.$$

Show that there exists a unique additive map $g: E \to F$ such that $|g - f|$ is bounded (i.e. $|g(x) - f(x)|$ is bounded as a function of x). Generalize to the bilinear case. [*Hint*: Let

$$g(x) = \lim_{n \to \infty} \frac{f(2^n x)}{2^n}.]$$

14. (Tate) Let S be a set and $f: S \to S$ a map of S into itself. Let $h: S \to \mathbf{R}$ be a real valued function. Assume that there exists a real number $d > 1$ such that $h \circ f - df$ is bounded. Show that there exists a unique function h_f such that $h_f - h$ is bounded, and $h_f \circ f = dh_f$. [*Hint*: Let $h_f(x) = \lim h(f^n(x))/d^n$.]

15. Define maps of degree > 2, from one module into another. [*Hint*: For degree 3, consider the expression

$$f(x + y + z) - f(x + y) - f(x + z) - f(y + z) + f(x) + f(y) + f(z).]$$

Generalize the statement proved for quadratic maps to these higher-degree maps, i.e. the uniqueness of the various multilinear maps entering into their definitions.

Alternating forms

16. Let E be a vector space over a field k and let g be a bilinear form on E. Assume that whenever $x, y \in E$ are such that $g(x, y) = 0$, then $g(y, x) = 0$. Show that g is symmetric or alternating.

17. Let E be a module over $\mathbf{Z}$. Assume that E is free, of dimension $n \geqq 1$, and let f be a bilinear alternating form on E. Show that there exists a basis $\{e_i\}$ $(i = 1, \ldots, n)$ and an integer r such that $2r \leqq n$,

$$e_1 \cdot e_2 = a_1, \qquad e_3 \cdot e_4 = a_2, \ \ldots, e_{2r-1} \cdot e_{2r} = a_r$$

where $a_1, \ldots, a_r \in \mathbf{Z}$, $a_i \neq 0$, and a_i divides a_{i+1} for $i = 1, \ldots, r - 1$ and finally $e_i \cdot e_j = 0$ for all other pairs of indices $i \leqq j$. Show that the ideals $\mathbf{Z}a_i$ are uniquely determined. [*Hint*: Consider the injective homomorphism $\varphi_f: E \to E^\vee$ of E into the

dual space over $\mathbf{Z}$, viewing $\varphi_f(E)$ as a free submodule of $E^\vee$.]. Generalize to principal rings when you know the basis theorem for modules over these rings.

Remark. A basis as in Exercise 18 is called a **symplectic basis**. For one use of such a basis, see the theory of theta functions, as in my *Introduction to Algebraic and Abelian Functions* (Second Edition, Springer Verlag), Chapter VI, §3.

18. Let E be a finite-dimensional vector space over the reals, and let $\langle\ ,\ \rangle$ be a symmetric positive definite form. Let Ω be a non-degenerate alternating form on E. Show that there exists a direct sum decomposition

$$E = E_1 \oplus E_2$$

having the following property. If $x, y \in E$ are written

$$x = (x_1, x_2) \quad \text{with} \quad x_1 \in E_1 \quad \text{and} \quad x_2 \in E_2,$$

$$y = (y_1, y_2) \quad \text{with} \quad y_1 \in E_1 \quad \text{and} \quad y_2 \in E_2,$$

then $\Omega(x, y) = \langle x_1, y_2 \rangle - \langle x_2, y_1 \rangle$. [*Hint*: Use Corollary 8.3, show that A is positive definite, and take its square root to transform the direct sum decomposition obtained in that corollary.]

19. Show that the pfaffian of an alternating $n \times n$ matrix is 0 when n is odd.

20. Prove all the properties for the pfaffian stated in Artin's *Geometric Algebra* (*Interscience*, 1957), p. 142.

The Witt group

21. Show explicitly how $W(k)$ is a homomorphic image of $WG(k)$.

22. Show that $WG(k)$ can be expressed as a homomorphic image of $\mathbf{Z}[k^*/k^{*2}]$ [*Hint*: Use the existence of orthogonal bases.]

23. Witt's theorem is still true for alternating forms. Prove it or look it up in Artin (ref. in Exercise 20).

$SL_n(\mathbf{R})$

There is a whole area of linear algebraic groups, giving rise to an extensive algebraic theory as well as the possibility of doing Fourier analysis on such groups. The group $SL_n(\mathbf{R})$ (or $SL_n(\mathbf{C})$) can serve as a prototype, and a number of basic facts can be easily verified. Some of them are listed below as exercises. Readers wanting to see solutions can look them up in [JoL 01], *Spherical Inversion on* $SL_n(\mathbf{R})$, Chapter I.

24. **Iwasawa decomposition**. We start with $GL_n(\mathbf{R})$. Let:

$G = GL_n(\mathbf{R})$;

$K =$ subgroup of real unitary $n \times n$ matrices;

$U =$ group of real unipotent upper triangular matrices, that is having components 1 on the diagonal, arbitrary above the diagonal, and 0 below the diagonal;

A = group of diagonal matrices with positive diagonal components.

Prove that the product map $U \times A \times K \to UAK \subset G$ is actually a bijection. This amounts to Gram–Schmidt orthogonalization. Prove the similar statement in the complex case, that is, for $G(\mathbf{C}) = GL_n(\mathbf{C})$, $K(\mathbf{C})$ = complex unitary group, $U(\mathbf{C})$ = complex unipotent upper triangular group, and A the same group of positive diagonal matrices as in the real case.

25. Let now $G = SL_n(\mathbf{R})$, and let K, A be the corresponding subgroups having determinant 1. Show that the product $U \times A \times K \to UAK$ again gives a bijection with G.

26. Let $\mathfrak{a}$ be the $\mathbf{R}$-vector space of real diagonal matrices with trace 0. Let $\mathfrak{a}^\vee$ be the dual space. Let α_i $(i = 1, \dots, n-1)$ be the functional defined on an element $H = \text{diag}(h_1, \dots, h_n)$ by $\alpha_i(H) = h_i - h_{i+1}$. (a) Show that $\{\alpha_1, \dots, \alpha_{n-1}\}$ is a basis of $\mathfrak{a}^\vee$ over $\mathbf{R}$. (b) Let $H_{i,i+1}$ be the diagonal matrix with $h_i = 1$, $h_{i+1} = -1$, and $h_j = 0$ for $j \neq i, i+1$. Show that $\{H_{1,2}, \dots, H_{n-1,n}\}$ is a basis of $\mathfrak{a}$. (c) Abbreviate $H_{i,i+1} = H_i$ $(i = 1, \dots, n-1)$. Let $\alpha'_i \in \mathfrak{a}^\vee$ be the functional such that $\alpha'_i(H_j) = \delta_{ij}$ ($= 1$ if $i = j$ and 0 otherwise). Thus $\{\alpha'_1, \dots, \alpha'_{n-1}\}$ is the dual basis of $\{H_1, \dots, H_{n-1}\}$. Show that

$$\alpha'_i(H) = h_1 + \cdots + h_i.$$

27. **The trace form.** Let $\text{Mat}_n(\mathbf{R})$ be the vector space of real $n \times n$ matrices. Define the **twisted trace form** on this space by

$$B_t(X, Y) = \text{tr}(X\,{}^tY) = \langle X, Y \rangle_t.$$

As usual, ${}^tY$ is the transpose of a matrix Y. Show that B_t is a symmetric positive definite bilinear form on $\text{Mat}_n(\mathbf{R})$. What is the analogous positive definite hermitian form on $\text{Mat}_n(\mathbf{C})$?

28. **Positivity.** On $\mathfrak{a}$ (real diagonal matrices with trace 0) the form of Exercise 27 can be defined by $\text{tr}(XY)$, since elements $X, Y \in \mathfrak{a}$ are symmetric. Let $\mathscr{A} = \{\alpha_1, \dots, \alpha_{n-1}\}$ denote the basis of Exercise 26. Define an element $H \in \mathfrak{a}$ to be **semipositive** (writen $H \geqq 0$) if $\alpha_i(H) \geqq 0$ for all $i = 1, \dots, n-1$. For each $\alpha \in \mathfrak{a}^\vee$, let $H_\alpha \in \mathfrak{a}$ represent α with respect to B_t, that is $\langle H_\alpha, H \rangle = \alpha(H)$ for all $H \in \mathfrak{a}$. Show that $H \geqq 0$ if and only if

$$H = \sum_{i=1}^{n-1} s_i H_{\alpha'_i} \qquad \text{with } s_i \geqq 0.$$

Similarly, define H to be **positive** and formulate the similar condition with $s_i > 0$.

29. Show that the elements $n\alpha'_i$ $(i = 1, \dots, n-1)$ can be expressed as linear combinations of $\alpha_1, \dots, \alpha_{n-1}$ with positive coefficients in $\mathbf{Z}$.

30. Let W be the group of permutations of the diagonal elements in the vector space $\mathfrak{a}$ of diagonal matrices. Show that $\mathfrak{a}_{\geqq 0}$ is a fundamental domain for the action of W on $\mathfrak{a}$ (i.e., given $H \in \mathfrak{a}$, there exists a unique $H^+ \geqq 0$ such that $H^+ = wH$ for some $w \in W$.

CHAPTER XVI

The Tensor Product

Having considered bilinear maps, we now come to multilinear maps and basic theorems concerning their structure. There is a universal module representing multilinear maps, called the tensor product. We derive its basic properties, and postpone to Chapter XIX the special case of alternating products. The tensor product derives its name from the use made in differential geometry, when this product is applied to the tangent space or cotangent space of a manifold. The tensor product can be viewed also as providing a mechanism for "extending the base"; that is, passing from a module over a ring to a module over some algebra over the ring. This "extension" can also involve reduction modulo an ideal, because what matters is that we are given a ring homomorphism $f: A \to B$, and we pass from modules over A to modules over B. The homomorphism f can be of both types, an inclusion or a canonical map with $B = A/J$ for some ideal J, or a composition of the two.

I have tried to provide the basic material which is immediately used in a variety of applications to many fields (topology, algebra, differential geometry, algebraic geometry, etc.).

§1. TENSOR PRODUCT

Let R be a commutative ring. If $E_1, \dots, E_n, F$ are modules, we denote by

$$L^n(E_1, \dots, E_n; F)$$

the module of n-multilinear maps

$$f: E_1 \times \cdots \times E_n \to F.$$

We recall that a multilinear map is a map which is linear (i.e., R-linear) in each variable. We use the words *linear* and *homomorphism* interchangeably. *Unless otherwise specified, modules, homomorphisms, linear, multilinear refer to the ring R.*

One may view the multilinear maps of a fixed set of modules $E_1, \dots, E_n$ as the objects of a category. Indeed, if

$$f: E_1 \times \cdots \times E_n \to F \quad \text{and} \quad g: E_1 \times \cdots \times E_n \to G$$

are multilinear, we define a morphism $f \to g$ to be a homomorphism $h: F \to G$ which makes the following diagram commutative:

$$\begin{array}{ccc} & & F \\ & \overset{f}{\nearrow} & \\ E_1 \times \cdots \times E_n & & \downarrow h \\ & \underset{g}{\searrow} & \\ & & G \end{array}$$

A universal object in this category is called a **tensor product** of $E_1, \dots, E_n$ (over R).

We shall now prove that a tensor product exists, and in fact construct one in a natural way. By abstract nonsense, we know of course that a tensor product is uniquely determined, up to a unique isomorphism.

Let M be the free module generated by the set of all n-tuples $(x_1, \dots, x_n)$, $(x_i \in E_i)$, i.e. generated by the set $E_1 \times \cdots \times E_n$. Let N be the submodule generated by all the elements of the following type:

$$(x_1, \dots, x_i + x_i', \dots, x_n) - (x_1, \dots, x_i, \dots, x_n) - (x_1, \dots, x_i', \dots, x_n)$$
$$(x_1, \dots, ax_i, \dots, x_n) - a(x_1, \dots, x_n)$$

for all $x_i \in E_i$, $x_i' \in E_i$, $a \in R$. We have the canonical injection

$$E_1 \times \cdots \times E_n \to M$$

of our set into the free module generated by it. We compose this map with the canonical map $M \to M/N$ on the factor module, to get a map

$$\varphi: E_1 \times \cdots \times E_n \to M/N.$$

We contend that φ is multilinear and is a tensor product.

It is obvious that φ is multilinear—our definition was adjusted to this purpose. Let

$$f: E_1 \times \cdots \times E_n \to G$$

be a multilinear map. By the definition of free module generated by

$$E_1 \times \cdots \times E_n$$

we have an induced linear map $M \to G$ which makes the following diagram commutative:

$$\begin{array}{ccc} & & M \\ & \nearrow & \downarrow \\ E_1 \times \cdots \times E_n & & \\ & \underset{f}{\searrow} & G \end{array}$$

Since f is multilinear, the induced map $M \to G$ takes on the value 0 on N. Hence by the universal property of factor modules, it can be factored through M/N, and we have a homomorphism $f_* : M/N \to G$ which makes the following diagram commutative:

$$\begin{array}{ccc} & & M/N \\ & \overset{\varphi}{\nearrow} & \downarrow f_* \\ E_1 \times \cdots \times E_n & & \\ & \underset{f}{\searrow} & G \end{array}$$

Since the image of φ generates M/N, it follows that the induced map f_* is uniquely determined. This proves what we wanted.

The module M/N will be denoted by

$$E_1 \otimes \cdots \otimes E_n \quad \text{or also} \quad \bigotimes_{i=1}^{n} E_i.$$

We have constructed a specific tensor product in the isomorphism class of tensor products, and we shall call it **the** tensor product of $E_1, \ldots, E_n$. If $x_i \in E_i$, we write

$$\varphi(x_1, \ldots, x_n) = x_1 \otimes \cdots \otimes x_n = x_1 \otimes_R \cdots \otimes_R x_n.$$

We have for all i,

$$x_1 \otimes \cdots \otimes ax_i \otimes \cdots \otimes x_n = a(x_1 \otimes \cdots \otimes x_n),$$

$$x_1 \otimes \cdots \otimes (x_i + x_i') \otimes \cdots \otimes x_n$$
$$= (x_1 \otimes \cdots \otimes x_n) + (x_1 \otimes \cdots \otimes x_i' \otimes \cdots \otimes x_n)$$

for $x_i, x_i' \in E_i$ and $a \in R$.

If we have two factors, say $E \otimes F$, then every element of $E \otimes F$ can be written as a sum of terms $x \otimes y$ with $x \in E$ and $y \in F$, because such terms generate $E \otimes F$ over k, and $a(x \otimes y) = ax \otimes y$ for $a \in R$.

Remark. If an element of the tensor product is 0, then that element can already be expressed in terms of a finite number of the relations defining the tensor product. Thus if E is a direct limit of submodules E_i then

$$\varinjlim F \otimes E_i = F \otimes \varinjlim E_i = F \otimes E.$$

In particular, every module is a direct limit of finitely generated submodules, and one uses frequently the technique of testing whether an element of $F \otimes E$ is 0 by testing whether the image of this element in $F \otimes E_i$ is 0 when E_i ranges over the finitely generated submodules of E.

Warning. The tensor product can involve a great deal of collapsing between the modules. For instance, take the tensor product over $\mathbf{Z}$ of $\mathbf{Z}/m\mathbf{Z}$ and $\mathbf{Z}/n\mathbf{Z}$ where m, n are integers > 1 and are relatively prime. Then the tensor product

$$\mathbf{Z}/n\mathbf{Z} \otimes \mathbf{Z}/m\mathbf{Z} = 0.$$

Indeed, we have $n(x \otimes y) = (nx) \otimes y = 0$ and $m(x \otimes y) = x \otimes my = 0$. Hence $x \otimes y = 0$ for all $x \in \mathbf{Z}/n\mathbf{Z}$ and $y \in \mathbf{Z}/m\mathbf{Z}$. Elements of type $x \otimes y$ generate the tensor product, which is therefore 0. We shall see later conditions under which there is no collapsing.

In many subsequent results, we shall assert the existence of certain linear maps from a tensor product. This existence is proved by using the universal mapping property of bilinear maps factoring through the tensor product. The uniqueness follows by prescribing the value of the linear maps on elements of type $x \otimes y$ (say for two factors) since such elements generate the tensor product.

We shall prove the associativity of the tensor product.

Proposition 1.1. *Let E_1, E_2, E_3 be modules. Then there exists a unique isomorphism*

$$(E_1 \otimes E_2) \otimes E_3 \to E_1 \otimes (E_2 \otimes E_3)$$

such that

$$(x \otimes y) \otimes z \mapsto x \otimes (y \otimes z)$$

for $x \in E_1$, $y \in E_2$ and $z \in E_3$.

Proof. Since elements of type $(x \otimes y) \otimes z$ generate the tensor product, the uniqueness of the desired linear map is obvious. To prove its existence, let $x \in E_1$. The map

$$\lambda_x : E_2 \times E_3 \to (E_1 \otimes E_2) \otimes E_3$$

such that $\lambda_x(y, z) = (x \otimes y) \otimes z$ is clearly bilinear, and hence factors through a linear map of the tensor product

$$\bar{\lambda}_x : E_2 \otimes E_3 \to (E_1 \otimes E_2) \otimes E_3.$$

The map

$$E_1 \times (E_2 \otimes E_3) \to (E_1 \otimes E_2) \otimes E_3$$

such that

$$(x, \alpha) \mapsto \bar{\lambda}_x(\alpha)$$

for $x \in E_1$ and $\alpha \in E_2 \otimes E_3$ is then obviously bilinear, and factors through a linear map

$$E_1 \otimes (E_2 \otimes E_3) \to (E_1 \otimes E_2) \otimes E_3,$$

which has the desired property (clear from its construction).

Proposition 1.2. *Let E, F be modules. Then there is a unique isomorphism*

$$E \otimes F \to F \otimes E$$

such that $x \otimes y \mapsto y \otimes x$ for $x \in E$ and $y \in F$.

Proof. The map $E \times F \to F \otimes E$ such that $(x, y) \mapsto y \otimes x$ is bilinear, and factors through the tensor product $E \otimes F$, sending $x \otimes y$ on $y \otimes x$. Since this last map has an inverse (by symmetry) we obtain the desired isomorphism.

The tensor product has various functorial properties. First, suppose that

$$f_i : E'_i \to E_i \qquad (i = 1, \dots, n)$$

is a collection of linear maps. We get an induced map on the product,

$$\prod f_i : \prod E'_i \to \prod E_i.$$

If we compose $\prod f_i$ with the canonical map into the tensor product, then we get an induced linear map which we may denote by $T(f_1, \dots, f_n)$ which makes the following diagram commutative:

$$\begin{array}{ccc} E'_1 \times \cdots \times E'_n & \longrightarrow & E'_1 \otimes \cdots \otimes E'_n \\ {\scriptstyle \prod f_i}\downarrow & & \downarrow{\scriptstyle T(f_1, \dots, f_n)} \\ E_1 \times \cdots \times E_n & \longrightarrow & E_1 \otimes \cdots \otimes E_n \end{array}$$

It is immediately verified that T is functorial, namely that if we have a composite of linear maps $f_i \circ g_i$ $(i = 1, \ldots, n)$ then

$$T(f_1 \circ g_1, \ldots, f_n \circ g_n) = T(f_1, \ldots, f_n) \circ T(g_1, \ldots, g_n)$$

and

$$T(\mathrm{id}, \ldots, \mathrm{id}) = \mathrm{id}.$$

We observe that $T(f_1, \ldots, f_n)$ is the unique linear map whose effect on an element $x'_1 \otimes \cdots \otimes x'_n$ of $E'_1 \otimes \cdots \otimes E'_n$ is

$$x'_1 \otimes \cdots \otimes x'_n \mapsto f_1(x'_1) \otimes \cdots \otimes f_n(x'_n).$$

We may view T as a map

$$\prod_{i=1}^{n} L(E'_i, E_i) \to L\left(\bigotimes_{i=1}^{n} E'_i, \bigotimes_{i=1}^{n} E_i\right),$$

and the reader will have no difficulty in verifying that this map is multilinear. We shall write out what this means explicitly for two factors, so that our map can be written

$$(f, g) \mapsto T(f, g).$$

Given homomorphisms $f : F' \to F$ and $g_1, g_2 : E' \to E$, then

$$T(f, g_1 + g_2) = T(f, g_1) + T(f, g_2),$$
$$T(f, ag_1) = aT(f, g_1).$$

In particular, select a fixed module F, and consider the functor $\tau = \tau_F$ (from modules to modules) such that

$$\tau(E) = F \otimes E.$$

Then τ gives rise to a linear map

$$\tau : L(E', E) \to L(\tau(E'), \tau(E))$$

for each pair of modules E', E, by the formula

$$\tau(f) = T(\mathrm{id}, f).$$

Remark. By abuse of notation, it is sometimes convenient to write

$$f_1 \otimes \cdots \otimes f_n \quad \text{instead of} \quad T(f_1, \ldots, f_n).$$

This should not be confused with the tensor product of elements taken in the tensor product of the modules

$$L(E'_1, E_1) \otimes \cdots \otimes L(E'_n, E_n).$$

The context will always make our meaning clear.

§2. BASIC PROPERTIES

The most basic relation relating linear maps, bilinear maps, and the tensor product is the following: For three modules E, F, G,

$$\boxed{L(E, L(F, G)) \approx L^2(E, F; G) \approx L(E \otimes F, G).}$$

The isomorphisms involved are described in a natural way.

(i) $L^2(E, F; G) \to L(E, L(F, G))$.

If $f : E \times F \to G$ is bilinear, and $x \in E$, then the map

$$f_x : F \to G$$

such that $f_x(y) = f(x, y)$ is linear. Furthermore, the map $x \mapsto f_x$ is linear, and is associated with f to get (i).

(ii) $L(E, L(F, G)) \to L^2(E, F; G)$.

Let $\varphi \in L(E, L(F, G))$. We let $f_\varphi : E \times F \to G$ be the bilinear map such that

$$f_\varphi(x, y) = \varphi(x)(y).$$

Then $\varphi \mapsto f_\varphi$ defines (ii).

It is clear that the homomorphisms of (i) and (ii) are inverse to each other and therefore give isomorphisms of the first two objects in the enclosed box.

(iii) $L^2(E, F; G) \to L(E \otimes F, G)$.

This is the map $f \mapsto f_*$ which associates to each bilinear map f the induced linear map on the tensor product. The association $f \mapsto f_*$ is injective (because f_* is uniquely determined by f), and it is surjective, because any linear map of the tensor product composed with the canonical map $E \times F \to E \otimes F$ gives rise to a bilinear map on $E \times F$.

Proposition 2.1. *Let $E = \bigoplus_{i=1}^{n} E_i$ be a direct sum. Then we have an isomorphism*

$$F \otimes E \leftrightarrow \bigoplus_{i=1}^{n} (F \otimes E_i).$$

Proof. The isomorphism is given by abstract nonsense. We keep F fixed, and consider the functor $\tau : X \mapsto F \otimes X$. As we saw above, τ is linear. We have projections $\pi_i : E \to E$ of E on E_i. Then

$$\pi_i \circ \pi_i = \pi_i, \qquad \pi_i \circ \pi_j = 0 \qquad \text{if} \quad i \neq j,$$

$$\sum_{i=1}^{n} \pi_i = \text{id}.$$

We apply the functor τ, and see that $\tau(\pi_i)$ satisfies the same relations, hence gives a direct sum decomposition of $\tau(E) = F \otimes E$. Note that $\tau(\pi_i) = \text{id} \otimes \pi_i$.

Corollary 2.2. *Let I be an indexing set, and $E = \bigoplus_{i \in I} E_i$. Then we have an isomorphism*

$$\left(\bigoplus_{i \in I} E_i\right) \otimes F \approx \bigoplus_{i \in I} (E_i \otimes F).$$

Proof. Let S be a finite subset of I. We have a sequence of maps

$$\left(\bigoplus_{i \in S} E_i\right) \times F \to \bigoplus_{i \in S} (E_i \otimes F) \to \bigoplus_{i \in I} (E_i \otimes F)$$

the first of which is bilinear, and the second is linear, induced by the inclusion of S in I. The first is the obvious map. If $S \subset S'$, then a trivial commutative diagram shows that the restriction of the map

$$\left(\bigoplus_{i \in S'} E_i\right) \times F \to \bigoplus_{i \in I} (E_i \otimes F)$$

induces our preceding map on the sum for $i \in S$. But we have an *injection*

$$\left(\bigoplus_{i \in S} E_i\right) \times F \to \left(\bigoplus_{i \in S'} E_i\right) \times F.$$

Hence by compatibility, we can define a bilinear map

$$\left(\bigoplus_{i \in I} E_i\right) \times F \to \bigoplus_{i \in I} (E_i \otimes F),$$

and consequently a linear map

$$\left(\bigoplus_{i\in I} E_i\right) \otimes F \to \bigoplus_{i\in I} (E_i \otimes F).$$

In a similar way, one defines a map in the opposite direction, and it is clear that these maps are inverse to each other, hence give an isomorphism.

Suppose now that E is free, of dimension 1 over R. Let $\{v\}$ be a basis, and consider $F \otimes E$. Every element of $F \otimes E$ can be written as a sum of terms $y \otimes av$ with $y \in F$ and $a \in R$. However, $y \otimes av = ay \otimes v$. In a sum of such terms, we can then use linearity on the left,

$$\sum_{i=1}^{n} (y_i \otimes v) = \left(\sum_{i=1}^{n} y_i\right) \otimes v, \qquad y_i \in F.$$

Hence every element is in fact of type $y \otimes v$ with some $y \in F$.

We have a bilinear map

$$F \times E \to F$$

such that $(y, av) \mapsto ay$, inducing a linear map

$$F \otimes E \mapsto F.$$

We also have a linear map $F \to F \otimes E$ given by $y \mapsto y \otimes v$. It is clear that these maps are inverse to each other, and hence that we have an isomorphism

$$F \otimes E \approx F.$$

Thus every element of $F \otimes E$ can be written *uniquely* in the form $y \otimes v$, $y \in F$.

Proposition 2.3. *Let E be free over R, with basis $\{v_i\}_{i\in I}$. Then every element of $F \otimes E$ has a unique expression of the form*

$$\sum_{i\in I} y_i \otimes v_i, \qquad y_i \in F$$

with almost all $y_i = 0$.

Proof. This follows at once from the discussion of the 1-dimensional case, and the corollary of Proposition 2.1.

Corollary 2.4. *Let E, F be free over R, with bases $\{v_i\}_{i\in I}$ and $\{w_j\}_{j\in J}$ respectively. Then $E \otimes F$ is free, with basis $\{v_i \otimes w_j\}$. We have*

$$\dim(E \otimes F) = (\dim E)(\dim F).$$

Proof. Immediate from the proposition.

We see that when E is free over R, then there is no collapsing in the tensor product. Every element of $F \otimes E$ can be viewed as a "formal" linear combination of elements in a basis of E with coefficients in F.

In particular, we see that $R \otimes E$ (or $E \otimes R$) is isomorphic to E, under the correspondence $x \mapsto x \otimes 1$.

Proposition 2.5. *Let E, F be free of finite dimension over R. Then we have an isomorphism*

$$\mathrm{End}_R(E) \otimes \mathrm{End}_R(F) \to \mathrm{End}_R(E \otimes F)$$

which is the unique linear map such that

$$f \otimes g \mapsto T(f, g)$$

for $f \in \mathrm{End}_R(E)$ and $g \in \mathrm{End}_R(F)$.

[We note that the tensor product on the left is here taken in the tensor product of the two modules $\mathrm{End}_R(E)$ and $\mathrm{End}_R(F)$.]

Proof. Let $\{v_i\}$ be a basis of E and let $\{w_j\}$ be a basis of F. Then $\{v_i \otimes w_j\}$ is a basis of $E \otimes F$. For each pair of indices (i', j') there exists a unique endomorphism $f = f_{i,i'}$ of E and $g = g_{j,j'}$ of F such that

$$f(v_i) = v_{i'} \quad \text{and} \quad f(v_\nu) = 0 \quad \text{if } \nu \neq i$$

$$g(w_j) = w_{j'} \quad \text{and} \quad g(w_\mu) = 0 \quad \text{if } \mu \neq j.$$

Furthermore, the families $\{f_{i,i'}\}$ and $\{g_{j,j'}\}$ are bases of $\mathrm{End}_R(E)$ and $\mathrm{End}_R(F)$ respectively. Then

$$T(f, g)(v_\nu \otimes w_\mu) = \begin{cases} v_{i'} \otimes w_{j'} & \text{if } (\nu, \mu) = (i, j) \\ 0 & \text{if } (\nu, \mu) \neq (i, j). \end{cases}$$

Thus the family $\{T(f_{i,i'}, g_{j,j'})\}$ is a basis of $\mathrm{End}_R(E \otimes F)$. Since the family $\{f_{i,i'} \otimes g_{j,j'}\}$ is a basis of $\mathrm{End}_R(E) \otimes \mathrm{End}_R(F)$, the assertion of our proposition is now clear.

In Proposition 2.5, we see that the ambiguity of the tensor sign in $f \otimes g$ is in fact unambiguous in the important special case of free, finite dimensional modules. We shall see later an important application of Proposition 2.5 when we discuss the tensor algebra of a module.

Proposition 2.6. *Let*

$$0 \to E' \xrightarrow{\varphi} E \xrightarrow{\psi} E'' \to 0$$

be an exact sequence, and F any module. Then the sequence

$$F \otimes E' \to F \otimes E \to F \otimes E'' \to 0$$

is exact.

Proof. Given $x'' \in E''$ and $y \in F$, there exists $x \in E$ such that $x'' = \psi(x)$, and hence $y \otimes x''$ is the image of $y \otimes x$ under the linear map

$$F \otimes E \to F \otimes E''.$$

Since elements of type $y \otimes x''$ generate $F \otimes E''$, we conclude that the preceding linear map is surjective. One also verifies trivially that the image of

$$F \otimes E' \to F \otimes E$$

is contained in the kernel of

$$F \otimes E \to F \otimes E''.$$

Conversely, let I be the image of $F \otimes E' \to F \otimes E$, and let

$$f : (F \otimes E)/I \to F \otimes E''$$

be the canonical map. We shall define a linear map

$$g : F \otimes E'' \to (F \otimes E)/I$$

such that $g \circ f = \text{id}$, This obviously will imply that f is injective, and hence will prove the desired converse.

Let $y \in F$ and $x'' \in E''$. Let $x \in E$ be such that $\psi(x) = x''$. We define a map $F \times E'' \to (F \otimes E)/I$ by letting

$$(y, x'') \mapsto y \otimes x \pmod{I},$$

and contend that this map is well defined, i.e. independent of the choice of x such that $\psi(x) = x''$. If $\psi(x_1) = \psi(x_2) = x''$, then $\psi(x_1 - x_2) = 0$, and by hypothesis, $x_1 - x_2 = \varphi(x')$ for some $x' \in E'$. Then

$$y \otimes x_1 - y \otimes x_2 = y \otimes (x_1 - x_2) = y \otimes \varphi(x').$$

This shows that $y \otimes x_1 \equiv y \otimes x_2 \pmod{I}$, and proves that our map is well defined. It is obviously bilinear, and hence factors through a linear map g, on the tensor product. It is clear that the restriction of $g \circ f$ on elements of type $y \otimes x''$ is the identity. Since these elements generate $F \otimes E''$, we conclude that f is injective, as was to be shown.

It is not always true that the sequence

$$0 \to F \otimes E' \to F \otimes E \to F \otimes E'' \to 0$$

is exact. It is exact if the first sequence in Proposition 2.6 splits, i.e. if E is essentially the direct sum of E' and E''. This is a trivial consequence of Proposition 2.1, and the reader should carry out the details to get accustomed to the formalism of the tensor product.

Proposition 2.7. *Let $\mathfrak{a}$ be an ideal of R. Let E be a module. Then the map $(R/\mathfrak{a}) \times E \to E/\mathfrak{a}E$ induced by*

$$(a, x) \mapsto ax \pmod{\mathfrak{a}E}, \qquad a \in R, x \in E$$

is bilinear and induces an isomorphism

$$(R/\mathfrak{a}) \otimes E \xrightarrow{\approx} E/\mathfrak{a}E.$$

Proof. Our map $(a, x) \mapsto ax \pmod{\mathfrak{a}E}$ clearly induces a bilinear map of $R/\mathfrak{a} \times E$ onto $E/\mathfrak{a}E$, and hence a linear map of $R/\mathfrak{a} \otimes E$ onto $E/\mathfrak{a}E$. We can construct an inverse, for we have a well-defined linear map

$$E \to R/\mathfrak{a} \otimes E$$

such that $x \mapsto \bar{1} \otimes x$ (where $\bar{1}$ is the residue class of 1 in $R/\mathfrak{a}$). It is clear that $\mathfrak{a}E$ is contained in the kernel of this last linear map, and thus that we obtain a homomorphism

$$E/\mathfrak{a}E \to R/\mathfrak{a} \otimes E,$$

which is immediately verified to be inverse to the homomorphism described in the statement of the proposition.

The association $E \mapsto E/\mathfrak{a}E \approx R/\mathfrak{a} \otimes E$ is often called a **reduction map**. In §4, we shall interpret this reduction map as an extension of the base.

§3. FLAT MODULES

The question under which conditions the left-hand arrow in Proposition 2.6 is an injection gives rise to the theory of those modules for which it is, and we follow Serre in calling them flat. Thus formally, the following conditions are equivalent, and define a **flat** module F, which should be called **tensor exact**.

F 1. For every exact sequence

$$E' \to E \to E''$$

the sequence

$$F \otimes E' \to F \otimes E \to F \otimes E''$$

is exact.

F 2. For every short exact sequence

$$0 \to E' \to E \to E'' \to 0$$

the sequence

$$0 \to F \otimes E' \to F \otimes E \to F \otimes E'' \to 0$$

is exact.

F 3. For every injection $0 \to E' \to E$ the sequence

$$0 \to F \otimes E' \to F \otimes E$$

is exact.

It is immediate that **F 1** implies **F 2** implies **F 3**. Finally, we see that **F 3** implies **F 1** by writing down the kernel and image of the map $E' \to E$ and applying **F 3**. We leave the details to the reader.

The following proposition gives tests for flatness, and also examples.

Proposition 3.1.

(i) *The ground ring is flat as module over itself.*

(ii) *Let $F = \bigoplus F_i$ be a direct sum. Then F is flat if and only if each F_i is flat.*

(iii) *A projective module is flat.*

The properties expressed in this proposition are basically categorical, cf. the comments on abstract nonsense at the end of the section. In another vein, we have the following tests having to do with localization.

Proposition 3.2.

(i) *Let S be a multiplicative subset of R. Then $S^{-1}R$ is flat over R.*

(ii) *A module M is flat over R if and only if the localization $M_{\mathfrak{p}}$ is flat over $R_{\mathfrak{p}}$ for each prime ideal $\mathfrak{p}$ of R.*

(iii) *Let R be a principal ring. A module F is flat if and only if F is torsion free.*

The proofs are simple, and will be left to the reader. More difficult tests for flatness will be proved below, however.

Examples of non-flatness. If R is an entire ring, and a module M over R has torsion, then M is not flat. (Prove this, which is immediate.)

There is another type of example which illustrates another bad phenomenon. Let R be some ring in a finite extension K of $\mathbf{Q}$, and such that R is a finite module over $\mathbf{Z}$ but not integrally closed. Let R' be its integral closure. Let $\mathfrak{p}$ be a maximal ideal of R and suppose that $\mathfrak{p}R'$ is contained in two distinct maximal ideals $\mathfrak{P}_1$ and $\mathfrak{P}_2$. Then it can be shown that R' is not flat over R, otherwise R' would be free over the local ring $R_{\mathfrak{p}}$, and the rank would have to be 1, thus precluding the possibility of the two primes $\mathfrak{P}_1$ and $\mathfrak{P}_2$. It is good practice for the reader actually to construct a numerical example of this situation. The same type of example can be constructed with a ring $R = k[x,y]$, where k is an algebraically closed field, even of characteristic 0, and x, y are related by an irreducible polynomial equation $f(x,y) = 0$ over k. We take R not integrally closed, such that its integral closure exhibits the same splitting of a prime $\mathfrak{p}$ of R into two primes. In each one of these similar cases, one says that there is a singularity at $\mathfrak{p}$.

As a third example, let R be the power series ring in more than one variable over a field k. Let $\mathfrak{m}$ be the maximal ideal. Then $\mathfrak{m}$ is not flat, because otherwise, by Theorem 3.8 below, $\mathfrak{m}$ would be free, and if $R = k[[x_1, \ldots, x_n]]$, then $x_1, \ldots, x_n$ would be a basis for $\mathfrak{m}$ over R, which is obviously not the case, since x_1, x_2 are linearly dependent over R when $n \geqq 2$. The same argument, of course, applies to any local ring R such that $\mathfrak{m}/\mathfrak{m}^2$ has dimension $\geqq 2$ over $R/\mathfrak{m}$.

Next we come to further criteria when a module is flat. For the proofs, we shall snake it all over the place. Cf. the remark at the end of the section.

Lemma 3.3. *Let F be flat, and suppose that*

$$0 \to N \to M \to F \to 0$$

is an exact sequence. Then for any E, we have an exact sequence

$$0 \to N \otimes E \to M \otimes E \to F \otimes E \to 0.$$

Proof. Represent E as a quotient of a flat L by an exact sequence

$$0 \to K \to L \to E \to 0.$$

Then we have the following exact and commutative diagram:

$$\begin{array}{ccccccccc}
 & & & & & & 0 & & \\
 & & & & & & \downarrow & & \\
 & & N \otimes K & \longrightarrow & M \otimes K & \longrightarrow & F \otimes K & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & N \otimes L & \longrightarrow & M \otimes L & \longrightarrow & F \otimes L & & \\
 & & \downarrow & & \downarrow & & & & \\
 & & N \otimes E & \longrightarrow & M \otimes E & & & & \\
 & & \downarrow & & \downarrow & & & & \\
 & & 0 & & 0 & & & &
\end{array}$$

The top right 0 comes by hypothesis that F is flat. The 0 on the left comes from the fact that L is flat. The snake lemma yields the exact sequence

$$0 \to N \otimes E \to M \otimes E$$

which proves the lemma.

Proposition 3.4. *Let*

$$0 \to F' \to F \to F'' \to 0$$

be an exact sequence, and assume that F'' is flat. Then F is flat if and only if F' is flat. More generally, let

$$0 \to F^0 \to F^1 \to \cdots \to F^n \to 0$$

be an exact sequence such that $F^1, \ldots, F^n$ are flat. Then F^0 is flat.

Proof. Let $0 \to E' \to E$ be an injection. We have an exact and commutative diagram:

$$\begin{array}{ccccccccc} & & & & & & 0 & & \\ & & & & & & \downarrow & & \\ 0 & \longrightarrow & F' \otimes E' & \longrightarrow & F \otimes E' & \longrightarrow & F'' \otimes E' & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & F' \otimes E & \longrightarrow & F \otimes E & \longrightarrow & F'' \otimes E & & \end{array}$$

The 0 on top is by hypothesis that F'' is flat, and the two zeros on the left are justified by Lemma 3.3. If F' is flat, then the first vertical map is an injection, and the snake lemma shows that F is flat. If F is flat, then the middle column is an injection. Then the two zeros on the left and the commutativity of the left square show that the map $F' \otimes E' \to F' \otimes E$ is an injection, so F' is flat. This proves the first statement.

The proof of the second statement is done by induction, introducing kernels and cokernels at each step as in dimension shifting, and apply the first statement at each step. This proves the proposition

To give flexibility in testing for flatness, the next two lemmas are useful, in relating the notion of flatness to a specific module. Namely, we say that F is ***E*-flat** or **flat for** E, if for every monomorphism

$$0 \to E' \to E$$

the tensored sequence

$$0 \to F \otimes E' \to F \otimes E$$

is also exact.

Lemma 3.5. *Assume that F is E-flat. Then F is also flat for every submodule and every quotient module of E.*

Proof. The submodule part is immediate because if $E_1' \subset E_2' \subset E$ are submodules, and $F \otimes E_1' \to F \otimes E$ is a monomorphism so is $F \otimes E_1' \to F \otimes E_2'$ since the composite map with $F \otimes E_2' \to F \otimes E$ is a monomorphism. The only question lies with a factor module. Suppose we have an exact sequence

$$0 \to N \to E \to M \to 0.$$

Let M' be a submodule of M and E' its inverse image in E. Then we have a

commutative diagram of exact sequences:

$$\begin{array}{ccccccccc}
0 & \longrightarrow & N & \longrightarrow & E' & \longrightarrow & M' & \longrightarrow & 0 \\
 & & \| & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & N & \longrightarrow & E & \longrightarrow & M & \longrightarrow & 0.
\end{array}$$

We tensor with F to get the exact and commutative diagram

$$\begin{array}{ccccccccc}
 & & & & 0 & & K & & \\
 & & & & \downarrow & & \downarrow & & \\
 & & F \otimes N & \longrightarrow & F \otimes E' & \longrightarrow & F \otimes M' & \longrightarrow & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & F \otimes N & \longrightarrow & F \otimes E & \longrightarrow & F \otimes M & & \\
 & & \downarrow & & & & & & \\
 & & 0 & & & & & &
\end{array}$$

where K is the questionable kernel which we want to prove is 0. But the snake lemma yields the exact sequence

$$0 \to K \to 0$$

which concludes the proof.

Lemma 3.6. *Let $\{E_i\}$ be a family of modules, and suppose that F is flat for each E_i. Then F is flat for their direct sum.*

Proof. Let $E = \bigoplus E_i$ be their direct sum. We have to prove that given any submodule E' of E, the sequence

$$0 \to F \otimes E' \to F \otimes E = \bigoplus F \otimes E_i$$

is exact. Note that if an element of $F \otimes E'$ becomes 0 when mapped into the direct sum, then it becomes 0 already in a finite subsum, so without loss of generality we may assume that the set of indices is finite. Then by induction, we can assume that the set of indices consists of two elements, so we have two modules E_1 and E_2, and $E = E_1 \oplus E_2$. Let N be a submodule of E. Let $N_1 = N \cap E_1$ and let N_2 be the image of N under the projection on E_2. Then

we have the following commutative and exact diagram:

$$\begin{array}{ccccccccc} & & 0 & & & & 0 & & \\ & & \downarrow & & & & \downarrow & & \\ & & N_1 & \longrightarrow & N & \longrightarrow & N_2 & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & E_1 & \longrightarrow & E & \longrightarrow & E_2 & & \end{array}$$

Tensoring with F we get the exact and commutative diagram:

$$\begin{array}{ccccccccc} & & 0 & & & & 0 & & \\ & & \downarrow & & & & \downarrow & & \\ & & F \otimes N_1 & \longrightarrow & F \otimes N & \longrightarrow & F \otimes N_2 & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & F \otimes E_1 & \longrightarrow & F \otimes E & \longrightarrow & F \otimes E_2 & & \end{array}$$

The lower left exactness is due to the fact that $E = E_1 \otimes E_2$. Then the snake lemma shows that the kernel of the middle vertical map is 0. This proves the lemma.

The next proposition shows that to test for flatness, it suffices to do so only for a special class of exact sequences arising from ideals.

Proposition 3.7. *F is flat if and only if for every ideal $\mathfrak{a}$ of R the natural map*

$$\mathfrak{a} \otimes F \to \mathfrak{a}F$$

is an isomorphism. In fact, F is flat if and only for every ideal $\mathfrak{a}$ of R tensoring the sequence

$$0 \to \mathfrak{a} \to R \to R/\mathfrak{a} \to 0$$

with F yields an exact sequence.

Proof. If F is flat, then tensoring with F and using Proposition 2.7 shows that the natural map is an isomorphism, because $\mathfrak{a}M$ is the kernel of $M \to M/\mathfrak{a}M$. Conversely, assume that this map is an isomorphism for all ideals $\mathfrak{a}$. This means

that F is R-flat. By Lemma 3.6 it follows that F is flat for an arbitrary direct sum of R with itself, and since any module M is a quotient of such a direct sum, Lemma 3.5 implies that F is M-flat, thus concluding the proof.

Remark on abstract nonsense. The proofs of Proposition 3.1(i), (ii), (iii), and Propositions 3.3 through 3.4 are basically rooted in abstract nonsense, and depend only on arrow theoretic arguments. Specifically, as in Chapter XX, §8, suppose that we have a bifunctor T on two distinct abelian categories $\mathfrak{A}$ and $\mathfrak{B}$ such that for each A, the functor $B \mapsto T(A, B)$ is right exact and for each B the functor $A \mapsto T(A, B)$ is right exact. Instead of "flat" we call an object A of $\mathfrak{A}$ t_T**-exact** if $B \mapsto T(A, B)$ is an exact functor; and we call an object L of $\mathfrak{B}$ T**-exact** if $A \mapsto T(A, L)$ is exact. Then the references to the base ring and free modules can be replaced by abstract nonsense conditions as follows.

In the use of L in Lemma 3.3, we need to assume that for every object E of B there is a $'T$-exact L and an epimorphism

$$L \to E \to 0.$$

For the analog of Proposition 3.7, we need to assume that there is some object R in $\mathfrak{B}$ for which F is R-exact, that is given an exact sequence

$$0 \to \mathfrak{a} \to R$$

then $0 \to T(F, \mathfrak{a}) \to T(F, R)$ is exact; and we also need to assume that R is a generator in the sense that every object B is the quotient of a direct sum of R with itself, then over some family of indices, and T respects direct sums.

The snake lemma is valid in arbitrary abelian categories, either because its proof is "functorial," or by using a representation functor to reduce it to the category of abelian groups. Take your pick.

In particular, we really don't need to have a commutative ring as base ring, this was done only for simplicity of language.

We now pass to somewhat different considerations.

Theorem 3.8. *Let R be a commutative local ring, and let M be a finite flat module over R. Then M is free. In fact, if $x_1, \dots, x_n \in M$ are elements of M whose residue classes are a basis of $M/\mathfrak{m}M$ over $R/\mathfrak{m}$, then $x_1, \dots, x_n$ form a basis of M over R.*

Proof. Let $R^{(n)} \to M$ be the map which sends the unit vectors of $R^{(n)}$ on $x_1, \dots, x_n$ respectively, and let N be its kernel. We get an exact sequence

$$0 \to N \to R^{(n)} \to M,$$

whence a commutative diagram

$$\begin{array}{ccccccc} & & \mathfrak{m} \otimes N & \longrightarrow & \mathfrak{m} \otimes R^{(n)} & \longrightarrow & \mathfrak{m} \otimes M \\ & & \downarrow f & & \downarrow g & & \downarrow h \\ 0 & \longrightarrow & N & \longrightarrow & R^{(n)} & \longrightarrow & M \end{array}$$

in which the rows are exact. Since M is assumed flat, the map h is an injection. By the snake lemma one gets an exact sequence

$$0 \to \text{coker } f \to \text{coker } g \to \text{coker } h,$$

and the arrow on the right is merely

$$R^{(n)}/\mathfrak{m}R^{(n)} \to M/\mathfrak{m}M,$$

which is an isomorphism by the assumption on $x_1, \ldots, x_n$. It follows that $\text{coker } f = 0$, whence $\mathfrak{m}N = N$, whence $N = 0$ by Nakayama if R is Noetherian, so N is finitely generated. If R is not assumed Noetherian, then one has to add a slight argument as follows in case M is finitely presented.

Lemma 3.9. *Assume that M is finitely presented, and let*

$$0 \to N \to E \to M \to 0$$

be exact, with E finite free. Then N is finitely generated.

Proof. Let

$$L_1 \to L_2 \to M \to 0$$

be a finite presentation of M, that is an exact sequence with L_1, L_2 finite free. Using the freeness, there exists a commutative diagram

$$\begin{array}{ccccccccc} & & L_1 & \longrightarrow & L_2 & \longrightarrow & M & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow \text{id} & & \\ 0 & \longrightarrow & N & \longrightarrow & E & \longrightarrow & M & \longrightarrow & 0 \end{array}$$

such that $L_2 \to E$ is surjective. Then the snake lemma gives at once the exact sequence

$$0 \to \text{coker}(L_1 \to N) \to 0,$$

so $\text{coker}(L_1 \to N) = 0$, whence N is an image of L_1 and is therefore finitely generated, thereby proving the lemma, and also completing the proof of Theorem 3.8 when M is finitely presented.

We still have not proved Theorem 3.8 in the fully general case. For this we use Matsumura's proof (see his *Commutative Algebra*, Chapter 2), based on the following lemma.

Lemma 3.10. *Assume that M is flat over R. Let $a_i \in A$, $x_i \in M$ for $i = 1, \dots, n$, and suppose that we have the relation*

$$\sum_{i=1}^{n} a_i x_i = 0.$$

Then there exists an integer s and elements $b_{ij} \in A$ and $y_j \in M$ $(j = 1, \dots, s)$ such that

$$\sum_i a_i b_{ij} = 0 \quad \textit{for all } j \quad \textit{and} \quad x_i = \sum_j b_{ij} y_j \quad \textit{for all } i.$$

Proof. We consider the exact sequence

$$0 \to K \to R^{(n)} \to R$$

where the map $R^{(n)} \to R$ is given by

$$(b_1, \dots, b_n) \mapsto \sum_{i=1}^{n} a_i b_i,$$

and K is its kernel. Since M is flat it follows that

$$K \otimes M \to M^{(n)} \xrightarrow{f_M} M$$

is exact, where f_M is given by

$$f_M(z_1, \dots, z_n) = \sum_{i=1}^{n} a_i z_i.$$

Therefore there exist elements $\beta_j \in K$ and $y_j \in M$ such that

$$(x_1, \dots, x_n) = \sum_{j=1}^{s} \beta_j y_j.$$

Write $\beta_j = (b_{1j}, \dots, b_{nj})$ with $b_{ij} \in R$. This proves the lemma.

We may now apply the lemma to prove the theorem in exactly the same way we proved that a finite projective module over a local ring is free in Chapter X, Theorem 4.4, by induction. This concludes the proof.

Remark. In the applications I know of, the base ring is Noetherian, and so one gets away with the very simple proof given at first. I did not want to obstruct the simplicity of this proof, and that is the reason I gave the additional technicalities in increasing order of generality.

Applications of homology. We end this section by pointing out a connection between the tensor product and the homological considerations of Chapter XX, §8 for those readers who want to pursue this trend of thoughts. The tensor product is a bifunctor to which we can apply the considerations of Chapter XX, §8. Let M, N be modules. Let

$$\cdots \to E_i \to E_{i-1} \to E_0 \to M \to 0$$

be a free or projective resolution of M, i.e. an exact sequence where E_i is free or projective for all $i \geqq 0$. We write this sequence as

$$E_M \to M \to 0.$$

Then by definition,

$$\text{Tor}_i(M, N) = i\text{-th homology of the complex } E \otimes N, \text{ that is of}$$

$$\cdots \to E_i \otimes N \to E_{i-1} \otimes N \to \cdots \to E_0 \otimes N \to 0.$$

This homology is determined up to a unique isomorphism. I leave to the reader to pick whatever convention is agreeable to fix one resolution to determine a fixed representation of $\text{Tor}_i(M, N)$, to which all others are isomorphic by a unique isomorphism.

Since we have a bifunctorial isomorphism $M \otimes N \approx N \otimes M$, we also get a bifunctorial isomorphism

$$\text{Tor}_i(M, N) \approx \text{Tor}_i(N, M)$$

for all i. See Propositions 8.2 and 8.2′ of Chapter XX.

Following general principles, we say that M has **Tor-dimension** $\leqq d$ if $\text{Tor}_i(M, N) = 0$ for all $i > d$ and all N. From Chapter XX, §8 we get the following result, which merely replaces T-exact by flat.

Theorem 3.11. *The following three conditions are equivalent concerning a module M.*

(i) *M is flat.*

(ii) $\text{Tor}_1(M, N) = 0$ *for all N.*

(iii) $\text{Tor}_i(M, N) = 0$ *for all $i \geqq 1$ and all N, in other words, M has* Tor-*dimension* 0.

Remark. Readers willing to use this characterization can replace some of the preceding proofs from 3.3 to 3.6 by a Tor-dimension argument, which is more formal, or at least formal in a different way, and may seem more rapid. The snake lemma was used ad hoc in each case to prove the desired result. The general homology theory simply replaces this use by the corresponding formal homological step, once the general theory of the derived functor has been carried out.

§4. EXTENSION OF THE BASE

Let R be a commutative ring and let E be a R-module. We specify R since we are going to work with several rings in a moment. Let $R \to R'$ be a homomorphism of commutative rings, so that R' is an R-algebra, and may be viewed as an R-module also. We have a 3-multilinear map

$$R' \times R' \times E \to R' \otimes E$$

defined by the rule

$$(a, b, x) \mapsto ab \otimes x.$$

This induces therefore a R-linear map

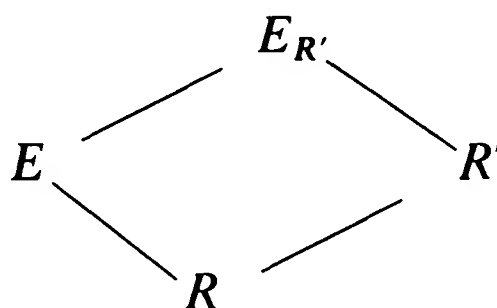
$$R' \otimes (R' \otimes E) \to R' \otimes E$$

and hence a R-bilinear map $R' \times (R' \otimes E) \to R' \otimes E$. It is immediately verified that our last map makes $R' \otimes E$ into a R'-module, which we shall call the **extension of** E **over** R', and denote by $E_{R'}$. We also say that $E_{R'}$ is obtained by **extension of the base** ring from R to R'.

Example 1. Let $\mathfrak{a}$ be an ideal of R and let $R \to R/\mathfrak{a}$ be the canonical homomorphism. Then the extension of E to $R/\mathfrak{a}$ is also called the **reduction** of E modulo $\mathfrak{a}$. This happens often over the integers, when we reduce modulo a prime p (i.e. modulo the prime ideal (p)).

Example 2. Let R be a field and R' an extension field. Then E is a vector space over R, and $E_{R'}$ is a vector space over R'. In terms of a basis, we see that our extension gives what was alluded to in the preceding chapter. This example will be expanded in the exercises.

We draw the same diagrams as in field theory:



to visualize an extension of the base. From Proposition 2.3, we conclude:

Proposition 4.1. *Let E be a free module over R, with basis $\{v_i\}_{i \in I}$. Let $v'_i = 1 \otimes v_i$. Then $E_{R'}$ is a free module over R', with basis $\{v'_i\}_{i \in I}$.*

We had already used a special case of this proposition when we observed that the dimension of a free module is defined, i.e. that two bases have the same

cardinality. Indeed, in that case, we reduced modulo a maximal ideal of R to reduce the question to a vector space over a field.

When we start changing rings, it is desirable to indicate R in the notation for the tensor product. Thus we write

$$E_{R'} = R' \otimes E = R' \otimes_R E.$$

Then we have transitivity of the extension of the base, namely, if $R \to R' \to R''$ is a succession of homomorphisms of commutative rings, then we have an isomorphism

$$R'' \otimes_R E \approx R'' \otimes_{R'} (R' \otimes_R E)$$

and this isomorphism is one of R''-modules. The proof is trivial and will be left to the reader.

If E has a multiplicative structure, we can extend the base also for this multiplication. Let $R \to A$ be a ring-homomorphism such that every element in the image of R in A commutes with every element in A (i.e. an R-algebra). Let $R \to R'$ be a homomorphism of commutative rings. We have a 4-multilinear map

$$R' \times A \times R' \times A \to R' \otimes A$$

defined by

$$(a, x, b, y) \mapsto ab \otimes xy.$$

We get an induced R-linear map

$$R' \otimes A \otimes R' \otimes A \to R' \otimes A$$

and hence an induced R-bilinear map

$$(R' \otimes A) \times (R' \otimes A) \to R' \otimes A.$$

It is trivially verified that the law of composition on $R' \otimes A$ we have just defined is associative. There is a unit element in $R' \otimes A$, namely, $1 \otimes 1$. We have a ring-homomorphism of R' into $R' \otimes A$, given by $a \mapsto a \otimes 1$. In this way one sees at once that $R' \otimes A = A_{R'}$ is an R'-algebra. We note that the map

$$x \mapsto 1 \otimes x$$

is a ring-homomorphism of A into $R' \otimes A$, and that we get a commutative diagram of ring homomorphisms,

$$\begin{array}{ccccc} & & R' \otimes A = A_{R'} & & \\ & \nearrow & & \nwarrow & \\ A & & & & R' \\ & \nwarrow & & \nearrow & \\ & & R & & \end{array}$$

For the record, we give some routine tests for flatness in the context of base extension.

Proposition 4.2. *Let $R \to A$ be an R-algebra, and assume A commutative.*

(i) **Base change.** *If F is a flat R-module, then $A \otimes_R F$ is a flat A-module.*

(ii) **Transitivity.** *If A is a flat commutative R-algebra and M is a flat A-module, then M is flat as R-module.*

The proofs are immediate, and will be left to the reader.

§5. SOME FUNCTORIAL ISOMORPHISMS

We recall an abstract definition. Let $\mathfrak{A}$, $\mathfrak{B}$ be two categories. The functors of $\mathfrak{A}$ into $\mathfrak{B}$ (say covariant, and in one variable) can be viewed as the objects of a category, whose morphisms are defined as follows. If L, M are two such functors, a morphism $H : L \to M$ is a rule which to each object X of $\mathfrak{A}$ associates a morphism $H_X : L(X) \to M(X)$ in $\mathfrak{B}$, such that for any morphism $f : X \to Y$ in $\mathfrak{A}$, the following diagram is commutative:

$$\begin{array}{ccc} L(X) & \xrightarrow{H_X} & M(X) \\ {\scriptstyle L(f)}\downarrow & & \downarrow{\scriptstyle M(f)} \\ L(Y) & \xrightarrow[H_Y]{} & M(Y) \end{array}$$

We can therefore speak of isomorphisms of functors. We shall see examples of these in the theory of tensor products below. In our applications, our categories are additive, that is, the set of morphisms is an additive group, and the composition law is $\mathbf{Z}$-bilinear. In that case, a functor L is called **additive** if

$$L(f + g) = L(f) + L(g).$$

We let R be a commutative ring, and we shall consider additive functors from the category of R-modules into itself. For instance we may view the dual module as a functor,

$$E \mapsto E^\vee = L(E, R) = \operatorname{Hom}_R(E, R).$$

Similarly, we have a functor in two variables,

$$(E, F) \mapsto L(E, F) = \operatorname{Hom}_R(E, F),$$

contravariant in the first, covariant in the second, and bi-additive.

We shall give several examples of functorial isomorphisms connected with the tensor product, and for this it is most convenient to state a general theorem, giving us a criterion when a morphism of functors is in fact an isomorphism.

Proposition 5.1. *Let L, M be two functors (both covariant or both contravariant) from the category of R-modules into itself. Assume that both functors are additive. Let $H : L \to M$ be a morphism of functors. If $H_E : L(E) \to M(E)$ is an isomorphism for every 1-dimensional free module E over R, then H_E is an isomorphism for every finite-dimensional free module over R.*

Proof. We begin with a lemma.

Lemma 5.2. *Let E and E_i $(i = 1, \dots, m)$ be modules over a ring. Let $\varphi_i : E_i \to E$ and $\psi_i : E \to E_i$ be homomorphisms having the following properties:*

$$\psi_i \circ \varphi_i = \text{id}, \qquad \psi_i \circ \varphi_j = 0 \quad \text{if} \quad i \neq j$$

$$\sum_{i=1}^{m} \varphi_i \circ \psi_i = \text{id},$$

Then the map

$$x \mapsto (\psi_1 x, \dots, \psi_m x)$$

is an isomorphism of E onto the direct product $\prod_{i=1}^{m} E_i$, and the map

$$(x_1, \dots, x_m) \mapsto \varphi_1 x_1 + \cdots + \varphi_m x_m$$

is an isomorphism of the product onto E. Conversely, if E is equal to the direct sum of submodules E_i $(i = 1, \dots, m)$, if we let ψ_i be the inclusion of E_i in E, and φ_i the projection of E on E_i, then these maps satisfy the above-mentioned properties.

Proof. The proof is routine, and is essentially the same as that of Proposition 3.1 of Chapter III. We shall leave it as an exercise to the reader.

We observe that the families $\{\varphi_i\}$ and $\{\psi_i\}$ satisfying the properties of the lemma behave functorially: If T is an additive contravariant functor, say, then the families $\{T(\psi_i)\}$ and $\{T(\varphi_i)\}$ also satisfy the properties of the lemma. Similarly if T is a covariant functor.

To apply the lemma, we take the modules E_i to be the 1-dimensional components occurring in a decomposition of E in terms of a basis. Let us assume for instance that L, M are both covariant. We have for each module E a com-

mutative diagram

$$\begin{array}{ccc} L(E) & \xrightarrow{H_E} & M(E) \\ {\scriptstyle L(\varphi_i)}\uparrow & & \uparrow{\scriptstyle M(\varphi_i)} \\ L(E_i) & \xrightarrow[H_{E_i}]{} & M(E_i) \end{array}$$

and a similar diagram replacing φ_i by ψ_i, reversing the two vertical arrows. Hence we get a direct sum decomposition of $L(E)$ in terms of $L(\psi_i)$ and $L(\varphi_i)$, and similarly for $M(E)$, in terms of $M(\psi_i)$ and $M(\varphi_i)$. By hypothesis, H_{E_i} is an isomorphism. It then follows trivially that H_E is an isomorphism. For instance, to prove injectivity, we write an element $v \in L(E)$ in the form

$$v = \sum L(\varphi_i)v_i,$$

with $v_i \in L(E_i)$. If $H_E v = 0$, then

$$0 = \sum H_E L(\varphi_i)v_i = \sum M(\varphi_i)H_{E_i}v_i,$$

and since the maps $M(\varphi_i)$ $(i = 1, \dots, m)$ give a direct sum decomposition of $M(E)$, we conclude that $H_{E_i}v_i = 0$ for all i, whence $v_i = 0$, and $v = 0$. The surjectivity is equally trivial.

When dealing with a functor of several variables, additive in each variable, one can keep all but one of the variables fixed, and then apply the proposition. We shall do this in the following corollaries.

Corollary 5.3. *Let E', E, F', F be free and finite dimensional over R. Then we have a functorial isomorphism*

$$L(E', E) \otimes L(F', F) \to L(E' \otimes F', E \otimes F)$$

such that

$$f \otimes g \mapsto T(f, g).$$

Proof. Keep E, F', F fixed, and view $L(E', E) \otimes L(F', F)$ as a functor in the variable E'. Similarly, view

$$L(E' \otimes F', E \otimes F)$$

as a functor in E'. The map $f \otimes g \mapsto T(f, g)$ is functorial, and thus by the lemma, it suffices to prove that it yields an isomorphism when E' has dimension 1. Assume now that this is the case; fix E' of dimension 1, and view the two expressions in the corollary as functors of the variable E. Applying the lemma

again, it suffices to prove that our arrow is an isomorphism when E has dimension 1. Similarly, we may assume that F, F' have dimension 1. In that case the verification that the arrow is an isomorphism is a triviality, as desired.

Corollary 5.4. *Let E, F be free and finite dimensional. Then we have a natural isomorphism*

$$\mathrm{End}_R(E) \otimes \mathrm{End}_R(F) \to \mathrm{End}_R(E \otimes F).$$

Proof. Special case of Corollary 5.3.

Note that Corollary 5.4 had already been proved before, and that we mention it here only to see how it fits with the present point of view.

Corollary 5.5. *Let E, F be free finite dimensional over R. There is a functorial isomorphism*

$$E^\vee \otimes F \to L(E, F)$$

given for $\lambda \in E^\vee$ and $y \in F$ by the map

$$\lambda \otimes y \mapsto A_{\lambda,y}$$

where $A_{\lambda,y}$ is such that for all $x \in E$, we have $A_{\lambda,y}(x) = \lambda(x)y$.

The inverse isomorphism of Corollary 5.5 can be described as follows. Let $\{v_1, \dots, v_n\}$ be a basis of E, and let $\{v_1^\vee, \dots, v_n^\vee\}$ be the dual basis. If $A \in L(E, F)$, then the element

$$\sum_{i=1}^{n} v_i^\vee \otimes A(v_i) \in E^\vee \otimes F$$

maps to A. In particular, if $E = F$, then the element mapping to the identity id_E is called the **Casimir element**

$$\sum_{i=1}^{n} v_i^\vee \otimes v_i,$$

independent of the choice of basis. Cf. Exercise 14.

To prove Corollary 5.5, justify that there is a well-defined homomorphism of $E^\vee \otimes F$ to $L(E, F)$, by the formula written down. Verify that this homomorphism is both injective and surjective. We leave the details as exercises.

Differential geometers are very fond of the isomorphism

$$L(E, E) \to E^\vee \otimes E,$$

and often use $E^\vee \otimes E$ when they think geometrically of $L(E, E)$, thereby emphasizing an unnecessary dualization, and an irrelevant formalism, when it is easier to deal directly with $L(E, E)$. In differential geometry, one applies various functors L to the tangent space at a point on a manifold, and elements of the spaces thus obtained are called **tensors** (of type L).

Corollary 5.6. *Let E, F be free and finite dimensional over R. There is a functorial isomorphism*

$$E^\vee \otimes F^\vee \to (E \otimes F)^\vee.$$

given for $x^\vee \in E^\vee$ and $y^\vee \in F^\vee$ by the map

$$x^\vee \otimes y^\vee \mapsto \lambda,$$

where λ is such that, for all $x \in E$ and $y \in F$,

$$\lambda(x \otimes y) = \langle x, x^\vee \rangle \langle y, y^\vee \rangle.$$

Proof. As before.

Finally, we leave the following results as an exercise.

Proposition 5.7. *Let E be free and finite dimensional over R. The trace function on $L(E, E)$ is equal to the composite of the two maps*

$$L(E, E) \to E^\vee \otimes E \to R,$$

where the first map is the inverse of the isomorphism described in Corollary 5.5, and the second map is induced by the bilinear map

$$(x^\vee, x) \mapsto \langle x, x^\vee \rangle.$$

Of course, it is precisely in a situation involving the trace that the isomorphism of Corollary 5.5 becomes important, and that the finite dimensionality of E is used. In many applications, this finite dimensionality plays no role, and it is better to deal with $L(E, E)$ directly.

§6. TENSOR PRODUCT OF ALGEBRAS

In this section, we again let R be a commutative ring. By an ***R*-algebra** we mean a ring homomorphism $R \to A$ into a ring A such that the image of R is contained in the center of A.

Let A, B be R-algebras. We shall make $A \otimes B$ into an R-algebra. Given $(a, b) \in A \times B$, we have an R-bilinear map

$$M_{a,b}: A \times B \to A \otimes B \text{ such that } M_{a,b}(a', b') = aa' \otimes bb'.$$

Hence $M_{a,b}$ induces an R-linear map $m_{a,b}: A \otimes B \to A \otimes B$ such that $m_{a,b}(a', b') = aa' \otimes bb'$. But $m_{a,b}$ depends bilinearly on a and b, so we obtain finally a unique R-bilinear map

$$A \otimes B \times A \otimes B \to A \otimes B$$

such that $(a \otimes b)(a' \otimes b') = aa' \otimes bb'$. This map is obviously associative, and we have a natural ring homomorphism

$$R \to A \otimes B \quad \text{given by} \quad c \mapsto 1 \otimes c = c \otimes 1.$$

Thus $A \otimes B$ is an R-algebra, called the **ordinary tensor product**.

Application: commutative rings

We shall now see the implication of the above for commutative rings.

Proposition 6.1. *Finite coproducts exist in the category of commutative rings, and in the category of commutative algebras over a commutative ring. If $R \to A$ and $R \to B$ are two homomorphisms of commutative rings, then their coproduct over R is the homomorphism $R \to A \otimes B$ given by*

$$a \mapsto a \otimes 1 = 1 \otimes a.$$

Proof. We shall limit our proof to the case of the coproduct of two ring homomorphisms $R \to A$ and $R \to B$. One can use induction.

Let A, B be commutative rings, and assume given ring-homomorphisms into a commutative ring C,

$$\varphi : A \to C \quad \text{and} \quad \psi : B \to C.$$

Then we can define a $\mathbf{Z}$-bilinear map

$$A \times B \to C$$

by $(x, y) \mapsto \varphi(x)\psi(y)$. From this we get a unique additive homomorphism

$$A \otimes B \to C$$

such that $x \otimes y \mapsto \varphi(x)\psi(y)$. We have seen above that we can define a ring structure on $A \otimes B$, such that

$$(a \otimes b)(c \otimes d) = ac \otimes bd.$$

It is then clear that our map $A \otimes B \to C$ is a ring-homomorphism. We also have two ring-homomorphisms

$$A \xrightarrow{f} A \otimes B \quad \text{and} \quad B \xrightarrow{g} A \otimes B$$

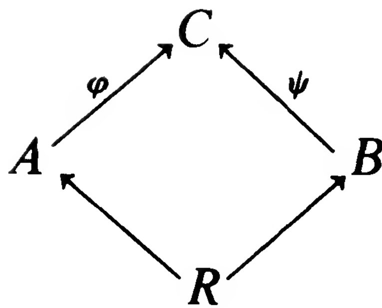
given by

$$x \mapsto x \otimes 1 \quad \text{and} \quad y \mapsto 1 \otimes y.$$

The universal property of the tensor product shows that $(A \otimes B, f, g)$ is a coproduct of our rings A and B.

If A, B, C are R-algebras, and if φ, ψ make the following diagram com-

mutative,



then $A \otimes B$ is also an R-algebra (it is in fact an algebra over R, or A, or B, depending on what one wants to use), and the map $A \otimes B \to C$ obtained above gives a homomorphism of R-algebras.

A commutative ring can always be viewed as a $\mathbf{Z}$-algebra (i.e. as an algebra over the integers). Thus one sees the coproduct of commutative rings as a special case of the coproduct of R-algebras.

Graded Algebras. Let G be a commutative monoid, written additively. By a ***G*-graded ring**, we shall mean a ring A, which as an additive group can be expressed as a direct sum.

$$A = \bigoplus_{r \in G} A_r,$$

and such that the ring multiplication maps $A_r \times A_s$ into A_{r+s}, for all $r, s \in G$.

In particular, we see that A_0 is a subring.

The elements of A_r are called the **homogeneous elements of degree *r***.

We shall construct several examples of graded rings, according to the following pattern. Suppose given for each $r \in G$ an abelian group A_r (written additively), and for each pair $r, s \in G$ a map $A_r \times A_s \to A_{r+s}$. Assume that A_0 is a commutative ring, and that composition under these maps is associative and A_0-bilinear. Then the direct sum $A = \bigoplus_{r \in G} A_r$ is a ring: We can define multiplication in the obvious way, namely

$$\left(\sum_{r \in G} x_r\right)\left(\sum_{s \in G} y_s\right) = \sum_{t \in G}\left(\sum_{r+s=t} x_r y_s\right).$$

The above product is called the **ordinary product**. However, there is another way. Suppose the grading is in $\mathbf{Z}$ or $\mathbf{Z}/2\mathbf{Z}$. We define the **super product** of $x \in A_r$ and $y \in A_s$ to be $(-1)^{rs}xy$, where xy is the given product. It is easily verified that this product is associative, and extends to what is called the **super product** $A \otimes A \to A$ associated with the bilinear maps. If R is a commutative ring such that A is a graded R-algebra, i.e. $RA_r \subset A_r$ for all r (in addition to the condition that A is a graded ring), then with the super product, A is also an R-algebra, which will be denoted by A_{su}, and will be called the **super algebra** associated with A.

Example. In the next section, we shall meet the tensor algebra $T(E)$, which will be graded as the direct sum of $T^r(E)$, and so we get the associated super tensor algebra $T_{su}(E)$ according to the above recipe.

Similarly, let A, B be graded algebras (graded by the natural numbers as above). We define their **super tensor product**

$$A \otimes_{su} B$$

to be the ordinary tensor product as graded module, but with the **super product**

$$(a \otimes b)(a' \otimes b') = (-1)^{(\deg b)(\deg a')} aa' \otimes bb'$$

if b, a' are homogeneous elements of B and A respectively. It is routinely verified that $A \otimes_{su} B$ is then a ring which is also a graded algebra. Except for the sign, the product is the same as the ordinary one, but it is necessary to verify associativity explicitly. Suppose $a' \in A_i$, $b \in B_j$, $a'' \in A_s$, and $b' \in B_r$. Then the reader will find at once that the sign which comes out by computing

$$(a \otimes_{su} b)(a' \otimes_{su} b')(a'' \otimes_{su} b'')$$

in two ways turns out to be the same, namely $(-1)^{ij+js+sr}$. Since bilinearity is trivially satisfied, it follows that $A \otimes_{su} B$ is indeed an algebra.

The super product in many ways is more natural than what we called the ordinary product. For instance, it is the natural product of cohomology in topology. Cf. Greenberg-Harper, *Algebraic Topology*, Chapter 29. For a similar construction with $\mathbf{Z}/2\mathbf{Z}$-grading, see Chapter XIX, §4.

§7. THE TENSOR ALGEBRA OF A MODULE

Let R be a commutative ring as before, and let E be a module (i.e. an R-module). For each integer $r \geqq 0$, we let

$$T^r(E) = \bigotimes_{i=1}^{r} E \quad \text{and} \quad T^0(E) = R.$$

Thus $T^r(E) = E \otimes \cdots \otimes E$ (tensor product taken r times). Then T^r is a functor, whose effect on linear maps is given as follows. If $f : E \to F$ is a linear map, then

$$T^r(f) = T(f, \ldots, f)$$

in the sense of §1.

From the associativity of the tensor product, we obtain a bilinear map

$$T^r(E) \times T^s(E) \to T^{r+s}(E),$$

which is associative. Consequently, by means of this bilinear map, we can define a ring structure on the direct sum

$$T(E) = \bigoplus_{r=0}^{\infty} T^r(E),$$

and in fact an algebra structure (mapping R on $T^0(E) = R$). We shall call $T(E)$ the **tensor algebra** of E, over R. It is in general *not* commutative. If $x, y \in T(E)$, we shall again write $x \otimes y$ for the ring operation in $T(E)$.

Let $f: E \to F$ be a linear map. Then f induces a linear map

$$T^r(f): T^r(E) \to T^r(F)$$

for each $r \geqq 0$, and in this way induces a map which we shall denote by $T(f)$ on $T(E)$. (There can be no ambiguity with the map of §1, which should now be written $T^1(f)$, and is in fact equal to f since $T^1(E) = E$.) It is clear that $T(f)$ is the unique linear map such that for $x_1, \ldots, x_r \in E$ we have

$$T(f)(x_1 \otimes \cdots \otimes x_r) = f(x_1) \otimes \cdots \otimes f(x_r).$$

Indeed, the elements of $T^1(E) = E$ are algebra-generators of $T(E)$ over R. We see that $T(f)$ is an algebra-homomorphism. *Thus T may be viewed as a functor from the category of modules to the category of graded algebras, $T(f)$ being a homomorphism of degree* 0.

When E is free and finite dimensional over R, we can determine the structure of $T(E)$ completely, using Proposition 2.3. Let P be an algebra over k. We shall say that P is a **non-commutative polynomial algebra** if there exist elements $t_1, \ldots, t_n \in P$ such that the elements

$$M_{(i)}(t) = t_{i_1} \cdots t_{i_s}$$

with $1 \leqq i_\nu \leqq n$ form a basis of P over R. We may call these elements non-commutative monomials in (t). As usual, by convention, when $r = 0$, the corresponding monomial is the unit element of P. We see that $t_1, \ldots, t_n$ generate P as an algebra over k, and that P is in fact a graded algebra, where P_r consists of linear combinations of monomials $t_{i_1} \cdots t_{i_r}$ with coefficients in R. It is natural to say that $t_1, \ldots, t_n$ are **independent non-commutative variables** over R.

Proposition 7.1. *Let E be free of dimension n over R. Then $T(E)$ is isomorphic to the non-commutative polynomial algebra on n variables over R. In other words, if $\{v_1, \ldots, v_n\}$ is a basis of E over R, then the elements*

$$M_{(i)}(v) = v_{i_1} \otimes \cdots \otimes v_{i_r}, \qquad 1 \leqq i_\nu \leqq n$$

form a basis of $T^r(E)$, and every element of $T(E)$ has a unique expression as a finite sum

$$\sum_{(i)} a_{(i)} M_{(i)}(v), \qquad a_{(i)} \in R$$

with almost all $a_{(i)}$ equal to 0.

Proof. This follows at once from Proposition 2.3.

The tensor product of linear maps will now be interpreted in the context of the tensor algebra.

For convenience, we shall denote the module of endomorphisms $\mathrm{End}_R(E)$ *by* $L(E)$ *for the rest of this section.*

We form the direct sum

$$(LT)(E) = \bigoplus_{r=0}^{\infty} L(T^r(E)),$$

which we shall also write $LT(E)$ for simplicity. (Of course, $LT(E)$ is not equal to $\mathrm{End}_R(T(E))$, so we must view LT as a single symbol.) We shall see that LT is a functor from modules to graded algebras, by defining a suitable multiplication on $LT(E)$. Let $f \in L(T^r(E))$, $g \in L(T^s(E))$, $h \in L(T^m(E))$. We define the product $fg \in L(T^{r+s}(E))$ to be $T(f, g)$, in the notation of §1, in other words to be the unique linear map whose effect on an element $x \otimes y$ with $x \in T^r(E)$ and $y \in T^s(E)$ is

$$x \otimes y \mapsto f(x) \otimes g(y).$$

In view of the associativity of the tensor product, we obtain at once the associativity $(fg)h = f(gh)$, and we also see that our product is bilinear. Hence $LT(E)$ is a k-algebra.

We have an algebra-homomorphism

$$T(L(E)) \to LT(E)$$

given in each dimension r by the linear map

$$f_1 \otimes \cdots \otimes f_r \mapsto T(f_1, \ldots, f_r) = f_1 \cdots f_r.$$

We specify here that the tensor product on the left is taken in

$$L(E) \otimes \cdots \otimes L(E).$$

We also note that the homomorphism is in general neither surjective nor injective. When E is free finite dimensional over R, the homomorphism turns out to be both, and thus we have a clear picture of $LT(E)$ as a non-commutative polynomial algebra, generated by $L(E)$. Namely, from Proposition 2.5, we obtain:

Proposition 7.2. *Let E be free, finite dimensional over R. Then we have an algebra-isomorphism*

$$T(L(E)) = T(\mathrm{End}_R(E)) \to LT(E) = \bigoplus_{r=0}^{\infty} \mathrm{End}_R(T^r(E))$$

given by

$$f \otimes g \mapsto T(f, g).$$

Proof. By Proposition 2.5, we have a linear isomorphism in each dimension, and it is clear that the map preserves multiplication.

In particular, we see that $LT(E)$ is a noncommutative polynomial algebra.

§8. SYMMETRIC PRODUCTS

Let $\mathfrak{S}_n$ denote the symmetric group on n letters, say operating on the integers $(1, \ldots, n)$. An r-multilinear map

$$f : E^{(r)} \to F$$

is said to be **symmetric** if $f(x_1, \ldots, x_r) = f(x_{\sigma(1)}, \ldots, x_{\sigma(r)})$ for all $\sigma \in \mathfrak{S}_r$.

In $T^r(E)$, we let $\mathfrak{b}_r$ be the submodule generated by all elements of type

$$x_1 \otimes \cdots \otimes x_r - x_{\sigma(1)} \otimes \cdots \otimes x_{\sigma(r)}$$

for all $x_i \in E$ and $\sigma \in \mathfrak{S}_r$. We define the factor module

$$S^r(E) = T^r(E)/\mathfrak{b}_r,$$

and let

$$S(E) = \bigoplus_{r=0}^{\infty} S^r(E)$$

be the direct sum. It is immediately obvious that the direct sum

$$\mathfrak{b} = \bigoplus_{r=0}^{\infty} \mathfrak{b}_r$$

is an ideal in $T(E)$, and hence that $S(E)$ is a graded R-algebra, which is called the **symmetric algebra** of E.

Furthermore, the canonical map

$$E^{(r)} \to S^r(E)$$

obtained by composing the maps

$$E^{(r)} \to T^r(E) \to T^r(E)/\mathfrak{b}_r = S^r(E)$$

is universal for r-multilinear symmetric maps.

We observe that S is a functor, from the category of modules to the category of graded R-algebras. The image of $(x_1, \ldots, x_r)$ under the canonical map

$$E^{(r)} \to S^r(E)$$

will be denoted simply by $x_1 \cdots x_r$.

Proposition 8.1. *Let E be free of dimension n over R. Let $\{v_1, \ldots, v_n\}$ be a basis of E over k. Viewed as elements of $S^1(E)$ in $S(E)$, these basis elements are algebraically independent over R, and $S(E)$ is therefore isomorphic to the polynomial algebra in n variables over R.*

Proof. Let $t_1, \ldots, t_n$ be algebraically independent variables over R, and form the polynomial algebra $R[t_1, \ldots, t_n]$. Let P_r be the R-module of homogeneous polynomials of degree r. We define a map of $E^{(r)} \to P_r$ as follows. If $w_1, \ldots, w_r$ are elements of E which can be written

$$w_i = \sum_{\nu=1}^{n} a_{i\nu} v_\nu, \qquad i = 1, \ldots, r,$$

then our map is given by

$$(w_1, \ldots, w_r) \mapsto (a_{11}t_1 + \cdots + a_{1n}t_n) \cdots (a_{r1}t_1 + \cdots + a_{rn}t_n).$$

It is obvious that this map is multilinear and symmetric. Hence it factors through a linear map of $S^r(E)$ into P_r:

$$\begin{array}{ccc} E^{(r)} & \longrightarrow & S^r(E) \\ & \searrow \quad \swarrow & \\ & P_r & \end{array}$$

From the commutativity of our diagram, it is clear that the element $v_{i_1} \cdots v_{i_s}$ in $S^r(E)$ maps on $t_{i_1} \cdots t_{i_s}$ in P_r for each r-tuple of integers $(i) = (i_1, \ldots, i_r)$. Since the monomials $M_{(i)}(t)$ of degree r are linearly independent over k, it follows that the monomials $M_{(i)}(v)$ in $S^r(E)$ are also linearly independent over R, and that our map $S^r(E) \to P_r$ is an isomorphism. One verifies at once that the multiplication in $S(E)$ corresponds to the multiplication of polynomials in $R[t]$, and thus that the map of $S(E)$ into the polynomial algebra described as above for each component $S^r(E)$ induces an algebra-isomorphism of $S(E)$ onto $R[t]$, as desired.

Proposition 8.2. *Let $E = E' \oplus E''$ be a direct sum of finite free modules. Then there is a natural isomorphism*

$$S^n(E' \oplus E'') \approx \bigoplus_{p+q=n} S^pE' \otimes S^qE''.$$

In fact, this is but the n-part of a graded isomorphism

$$S(E' \oplus E'') \approx SE' \otimes SE''.$$

Proof. The isomorphism comes from the following maps. The inclusions of E' and E'' into their direct sum give rise to the functorial maps

$$SE' \otimes SE'' \to SE,$$

and the claim is that this is a graded isomorphism. Note that SE' and SE'' are commutative rings, and so their tensor product is just the tensor product of commutative rings discussed in §6. The reader can either give a functorial map backward to prove the desired isomorphism, or more concretely, SE' is the polynomial ring on a finite family of variables, SE'' is the polynomial ring in another family of variables, and their tensor product is just the polynomial ring in the two families of variables. The matter is easy no matter what, and the formal proof is left to the reader.

EXERCISES

1. Let k be a field and $k(\alpha)$ a finite extension. Let $f(X) = \operatorname{Irr}(\alpha, k, X)$, and suppose that f is separable. Let k' be any extension of k. Show that $k(\alpha) \otimes k'$ is a direct sum of fields. If k' is algebraically closed, show that these fields correspond to the embeddings of $k(\alpha)$ in k'.
2. Let k be a field, $f(X)$ an irreducible polynomial over k, and α a root of f. Show that $k(\alpha) \otimes k'$ is isomorphic, as a k'-algebra, to $k'[X]/(f(X))$.
3. Let E be a finite extension of a field k. Show that E is separable over k if and only if $E \otimes_k L$ has no nilpotent elements for all extensions L of k, and also when $L = k^{a}$.
4. Let $\varphi : A \to B$ be a commutative ring homomorphism. Let E be an A-module and F a B-module. Let F_A be the A-module obtained from F via the operation of A on F through φ, that is for $y \in F_A$ and $a \in A$ this operation is given by

$$(a, y) \mapsto \varphi(a)y.$$

Show that there is a natural isomorphism

$$\operatorname{Hom}_B(B \otimes_A E, F) \approx \operatorname{Hom}_A(E, F_A).$$

5. **The norm.** Let B be a commutative algebra over the commutative ring R and assume that B is free of rank r. Let A be any commutative R-algebra. Then $A \otimes B$ is both an A-algebra and a B-algebra. We view $A \otimes B$ as an A-algebra, which is also free of rank r. If $\{e_1, \dots, e_r\}$ is a basis of B over R, then

$$1_A \otimes e_1, \dots, 1_A \otimes e_r$$

is a basis of $A \otimes B$ over A. We may then define the **norm**

$$N = N_{A \otimes B, A} : A \otimes B \to A$$

as the unique map which coincides with the determinant of the regular representation.

In other words, if $b \in B$ and b_B denotes multiplication by b, then

$$N_{B,R}(b) = \det(b_B);$$

and similarly after extension of the base. Prove:

(a) Let $\varphi : A \to C$ be a homomorphism of R-algebras. Then the following diagram is commutative:

$$\begin{array}{ccc} A \otimes B & \xrightarrow{\varphi \otimes \mathrm{id}} & C \otimes B \\ \downarrow N & & \downarrow N \\ A & \xrightarrow[\varphi]{} & C \end{array}$$

(b) Let $x, y \in A \otimes B$. Then $N(x \otimes_B y) = N(x) \otimes N(y)$. [*Hint*: Use the commutativity relations $e_i e_j = e_j e_i$ and the associativity.]

A little flatness

6. Let M, N be flat. Show that $M \otimes N$ is flat.
7. Let F be a flat R-module, and let $a \in R$ be an element which is not a zero-divisor. Show that if $ax = 0$ for some $x \in F$ then $x = 0$.
8. Prove Proposition 3.2.

Faithfully flat

9. We continue to assume that rings are commutative. Let M be an A-module. We say that M is **faithfully flat** if M is flat, and if the functor

$$T_M : E \mapsto M \otimes_A E.$$

is faithful, that is $E \neq 0$ implies $M \otimes_A E \neq 0$. Prove that the following conditions are equivalent.

(i) M is faithfully flat.

(ii) M is flat, and if $u : F \to E$ is a homomorphism of A-modules, $u \neq 0$, then $T_M(u) : M \otimes_A F \to M \otimes_A E$ is also $\neq 0$.

(iii) M is flat, and for all maximal ideals $\mathfrak{m}$ of A, we have $\mathfrak{m}M \neq M$.

(iv) A sequence of A-modules $N' \to N \to N''$ is exact if and only if the sequence tensored with M is exact.

10. (a) Let $A \to B$ be a ring-homomorphism. If M is faithfully flat over A, then $B \otimes_A M$ is faithfully flat over B.

(b) Let M be faithfully flat over B. Then M viewed as A-module via the homomorphism $A \to B$ is faithfully flat over A if B is faithfully flat over A.

11. Let P, M, E be modules over the commutative ring A. If P is finitely generated (resp. finitely presented) and E is flat, show that the natural homomorphism

$$\mathrm{Hom}_A(P, M) \otimes_A E \to \mathrm{Hom}_A(P, M \otimes_A E)$$

is a monomorphism (resp. an isomorphism).

[*Hint*: Let $F_1 \to F_0 \to P \to 0$ be a finite presentation, say. Consider the diagram

$$\begin{array}{ccccccc} 0 \longrightarrow & \mathrm{Hom}_A(P, M) \otimes_A E & \longrightarrow & \mathrm{Hom}_A(F_0, M) \otimes_A E & \longrightarrow & \mathrm{Hom}_A(F_1, M) \otimes_A E \\ & \downarrow & & \downarrow & & \downarrow \\ 0 \longrightarrow & \mathrm{Hom}_A(P, M \otimes_A E) & \longrightarrow & \mathrm{Hom}_A(F_0, M \otimes_A E) & \longrightarrow & \mathrm{Hom}_A(F_1, M \otimes_A E)]. \end{array}$$

Tensor products and direct limits

12. Show that the tensor product commutes with direct limits. In other words, if $\{E_i\}$ is a directed family of modules, and M is any module, then there is a natural isomorphism

$$\varinjlim (E_i \otimes_A M) \approx (\varinjlim E_i) \otimes_A M.$$

13. (D. Lazard) Let E be a module over a commutative ring A. Tensor products are all taken over that ring. Show that the following conditions are equivalent:
   (i) There exists a direct family $\{F_i\}$ of free modules of finite type such that
   $$E \approx \varinjlim F_i.$$
   (ii) E is flat.
   (iii) For every finitely presented module P the natural homomorphism
   $$\mathrm{Hom}_A(P, A) \otimes_A E \to \mathrm{Hom}_A(P, E)$$
   is surjective.
   (iv) For every finitely presented module P and homomorphism $f: P \to E$ there exists a free module F, finitely generated, and homomorphisms
   $$g: P \to F \quad \text{and} \quad h: F \to E$$
   such that $f = h \circ g$.

Remark. The point of Lazard's theorem lies in the first two conditions: *E is flat if and only if E is a direct limit of free modules of finite type.*

[*Hint*: Since the tensor product commutes with direct limits, that (i) implies (ii) comes from the preceding exercise and the definition of flat.

To show that (ii) implies (iii), use Exercise 11.

To show that (iii) implies (iv) is easy from the hypothesis.

To show that (iv) implies (i), use the fact that a module is a direct limit of finitely presented modules (an exercise in Chapter III), and (iv) to get the free modules instead. For complete details, see for instance Bourbaki, *Algèbre*, Chapter X, §1, Theorem 1, p. 14.]

The Casimir element

14. Let k be a commutative field and let E be a vector space over k, of finite dimension n. Let B be a nondegenerate symmetric bilinear form on E, inducing an iso-

morphism $E \to E^\vee$ of E with its dual space. Let $\{v_1, \dots, v_n\}$ be a basis of E. The B-dual basis $\{v'_1, \dots, v'_n\}$ consists of the elements of E such that $B(v_i, v'_j) = \delta_{ij}$.

(a) Show that the element $\sum v_i \otimes v'_i$ in $E \otimes E$ is independent of the choice of basis. We call this element the **Casimir element** (see below).

(b) In the symmetric algebra $S(E)$, let $Q_B = \sum v_i v'_i$. Show that Q_B is independent of the choice of basis. We call Q_B the **Casimir polynomial**. It depends on B, of course.

(c) More generally, let $\mathbf{D}$ be an (associative) algebra over k, let $\mathcal{D}: E \to \mathbf{D}$ be an injective linear map of E into $\mathbf{D}$. Show that the element $\sum \mathcal{D}(v_i)\mathcal{D}(v'_i) = \omega_{B,\mathcal{D}}$ is independent of the choice of basis. We call it the **Casimir element** in $\mathbf{D}$, determined by $\mathcal{D}$ and B.

Remark. The terminology of the Casimir element is determined by the classical case, when G is a Lie group, $E = \mathfrak{g} = \text{Lie}(G)$ is the Lie algebra of G (tangent space at the origin with the Lie algebra product determined by the Lie derivative), and $\mathcal{D}(v)$ is the differential operator associated with v (Lie derivative in the direction of v). The Casimir element is then a partial differential operator in the algebra of all differential operators on G. Cf. basic books on manifolds and Lie theory, for instance [JoL 01], Chapter II, §1 and Chapter VII, §2.

15. Let $E = \mathfrak{sl}_n(k) =$ subspace of $\text{Mat}_n(k)$ consisting of matrices with trace 0. Let B be the bilinear form defined by $B(X, Y) = \text{tr}(XY)$. Let $G = SL_n(k)$. Prove:

(a) B is $\mathbf{c}(G)$-invariant, where $\mathbf{c}(g)$ is conjugation by an element $g \in G$.

(b) B is invariant under the transpose $(X, Y) \mapsto ({}^tX, {}^tY)$.

(c) Let $k = \mathbf{R}$. Then B is positive definite on the symmetric matrices and negative definite on the skew-symmetric matrices.

(d) Suppose G is given with an action on the algebra $\mathbf{D}$ of Exercise 14, and that the linear map $\mathcal{D}: E \to \mathbf{D}$ is G-linear. Show that the Casimir element is G-invariant (for the conjugation action on $S(E)$, and the given action on $\mathbf{D}$).

CHAPTER XVII

Semisimplicity

In many applications, a module decomposes as a direct sum of simple submodules, and then one can develop a fairly precise structure theory, both under general assumptions, and particular applications. This chapter is devoted to those results which can be proved in general. In the next chapter, we consider those additional results which can be proved in a classical and important special case.

I have more or less followed Bourbaki in the proof of Jacobson's density theorem.

§1. MATRICES AND LINEAR MAPS OVER NON-COMMUTATIVE RINGS

In Chapter XIII, we considered exclusively matrices over commutative rings. For our present purposes, it is necessary to consider a more general situation.

Let K be a ring. We define a matrix (φ_{ij}) with coefficients in K just as we did for commutative rings. The product of matrices is defined by the same formula. Then we again have associativity and distributivity, whenever the size of the matrices involved in the operations makes the operations defined. In particular, the square $n \times n$ matrices over K form a ring, again denoted by $\mathrm{Mat}_n(K)$. We have a ring-homomorphism

$$K \to \mathrm{Mat}_n(K)$$

on the diagonal.

By a **division ring** we shall mean a ring with $1 \neq 0$, and such that every non-zero element has a multiplicative inverse.

If K is a division ring, then every non-zero K-module has a basis, and the cardinalities of two bases are equal. The proof is the same as in the commutative case; we never needed commutativity in the arguments. This cardinality is again called the dimension of the module over K, and a module over a division ring is called a vector space.

We can associate a matrix with linear maps, depending on the choice of a finite basis, just as in the commutative case. However, we shall consider a somewhat different situation which we want to apply to semisimple modules.

Let R be a ring, and let

$$E = E_1 \oplus \cdots \oplus E_n, \qquad F = F_1 \oplus \cdots \oplus F_m$$

be R-modules, expressed as direct sums of R-submodules. We wish to describe the most general R-homomorphism of E into F.

Suppose first $F = F_1$ has one component. Let

$$\varphi : E_1 \oplus \cdots \oplus E_n \to F$$

be a homomorphism. Let $\varphi_j : E_j \to F$ be the restriction of φ to the factor E_j. Every element $x \in E$ has a unique expression $x = x_1 + \cdots + x_n$, with $x_j \in E_j$. We may therefore associate with x the column vector $X = {}^t(x_1, \ldots, x_n)$, whose components are in $E_1, \ldots, E_n$ respectively. We can associate with φ the row vector $(\varphi_1, \ldots, \varphi_n)$, $\varphi_j \in \operatorname{Hom}_R(E_j, F)$, and the effect of φ on the element x of E is described by matrix multiplication, of the row vector times the column vector.

More generally, consider a homomorphism

$$\varphi : E_1 \oplus \cdots \oplus E_n \to F_1 \oplus \cdots \oplus F_m.$$

Let $\pi_i : F_1 \oplus \cdots \oplus F_m \to F_i$ be the projection on the i-th factor. Then we can apply our previous remarks to $\pi_i \circ \varphi$, for each i. In this way, we see that there exist unique elements $\varphi_{ij} \in \operatorname{Hom}_R(E_j, F_i)$, such that φ has a matrix representation

$$M(\varphi) = \begin{pmatrix} \varphi_{11} & \cdots & \varphi_{1n} \\ \vdots & & \vdots \\ \varphi_{m1} & \cdots & \varphi_{mn} \end{pmatrix}$$

whose effect on an element x is given by matrix multiplication, namely

$$\begin{pmatrix} \varphi_{11} & \cdots & \varphi_{1n} \\ \vdots & & \vdots \\ \varphi_{m1} & \cdots & \varphi_{mn} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

Conversely, given a matrix (φ_{ij}) with $\varphi_{ij} \in \mathrm{Hom}_R(E_j, F_i)$, we can define an element of $\mathrm{Hom}_R(E, F)$ by means of this matrix. We have an additive group-isomorphism between $\mathrm{Hom}_R(E, F)$ and this group of matrices.

In particular, let E be a fixed R-module, and let $K = \mathrm{End}_R(E)$. Then we have a ring-isomorphism

$$\mathrm{End}_R(E^{(n)}) \to \mathrm{Mat}_n(K)$$

which to each $\varphi \in \mathrm{End}_R(E^{(n)})$ associates the matrix

$$\begin{pmatrix} \varphi_{11} & \cdots & \varphi_{1n} \\ \vdots & & \vdots \\ \varphi_{n1} & \cdots & \varphi_{nn} \end{pmatrix}$$

determined as before, and operating on the left on column vectors of $E^{(n)}$, with components in E.

Remark. Let E be a 1-dimensional vector space over a division ring D, and let $\{v\}$ be a basis. For each $a \in D$, there exists a unique D-linear map $f_a : E \to E$ such that $f_a(v) = av$. Then we have the rule

$$f_a f_b = f_{ba}.$$

Thus when we associate a matrix with a linear map, depending on a basis, the multiplication gets twisted. Nevertheless, the statement we just made preceding this remark is correct!! The point is that we took the φ_{ij} in $\mathrm{End}_R(E)$, and not in D, in the special case that $R = D$. Thus K is not isomorphic to D (in the non-commutative case), but anti-isomorphic. This is the only point of difference of the formal elementary theory of linear maps in the commutative or non-commutative case.

We recall that an R-module E is said to be **simple** if it is $\neq 0$ and if it has no submodule other than 0 or E.

Proposition 1.1. Schur's Lemma. *Let E, F be simple R-modules. Every non-zero homomorphism of E into F is an isomorphism. The ring $\mathrm{End}_R(E)$ is a division ring.*

Proof. Let $f : E \to F$ be a non-zero homomorphism. Its image and kernel are submodules, hence $\mathrm{Ker}\, f = 0$ and $\mathrm{Im}\, f = F$. Hence f is an isomorphism. If $E = F$, then f has an inverse, as desired.

The next proposition describes completely the ring of endomorphisms of a direct sum of simple modules.

Proposition 1.2. *Let $E = E_1^{(n_1)} \oplus \cdots \oplus E_r^{(n_r)}$ be a direct sum of simple modules, the E_i being non-isomorphic, and each E_i being repeated n_i times in*

the sum. Then, up to a permutation, $E_1, \dots, E_r$ are uniquely determined up to isomorphisms, and the multiplicities $n_1, \dots, n_r$ are uniquely determined. The ring $\mathrm{End}_R(E)$ *is isomorphic to a ring of matrices, of type*

$$\begin{pmatrix} M_1 & & \cdots & 0 \\ \vdots & M_2 & & \vdots \\ & & \ddots & \\ 0 & & \cdots & M_r \end{pmatrix}$$

where M_i is an $n_i \times n_i$ matrix over $\mathrm{End}_R(E_i)$. *(The isomorphism is the one with respect to our direct sum decomposition.)*

Proof. The last statement follows from our previous considerations, taking into account Proposition 1.1.

Suppose now that we have two R-modules, with direct sum decompositions into simple submodules, and an isomorphism

$$E_1^{(n_1)} \oplus \cdots \oplus E_r^{(n_r)} \to F_1^{(m_1)} \oplus \cdots \oplus F_s^{(m_s)},$$

such that the E_i are non-isomorphic, and the F_j are non-isomorphic. From Proposition 1.1, we conclude that each E_i is isomorphic to some F_j, and conversely. It follows that $r = s$, and that after a permutation, $E_i \approx F_i$. Furthermore, the isomorphism must induce an isomorphism

$$E_i^{(n_i)} \to F_i^{(m_i)}$$

for each i. Since $E_i \approx F_i$, we may assume without loss of generality that in fact $E_i = F_i$. Thus we are reduced to proving: If a module is isomorphic to $E^{(n)}$ and to $E^{(m)}$, with some simple module E, then $n = m$. But $\mathrm{End}_R(E^{(n)})$ is isomorphic to the $n \times n$ matrix ring over the division ring $\mathrm{End}_R(E) = K$. Furthermore this isomorphism is verified at once to be an isomorphism as K-vector space. The dimension of the space of $n \times n$ matrices over K is n^2. This proves that the multiplicity n is uniquely determined, and proves our proposition.

When E admits a (finite) direct sum decomposition of simple submodules, the number of times that a simple module of a given isomorphism class occurs in a decomposition will be called the **multiplicity** of the simple module (or of the isomorphism class of the simple module).

Furthermore, if

$$E = E_1^{(n_1)} \oplus \cdots \oplus E_r^{(n_r)}$$

is expressed as a sum of simple submodules, we shall call $n_1 + \cdots + n_r$ the **length** of E. In many applications, we shall also write

$$E = n_1 E_1 \oplus \cdots \oplus n_r E_r = \bigoplus_{i=1}^{r} n_i E_i.$$

§2. CONDITIONS DEFINING SEMISIMPLICITY

Let R be a ring. Unless otherwise specified in this section all modules and homomorphisms will be R-modules and R-homomorphisms.

The following conditions on a module E are equivalent:

SS 1. E is the sum of a family of simple submodules.

SS 2. E is the direct sum of a family of simple submodules.

SS 3. Every submodule F of E is a direct summand, i.e. there exists a submodule F' such that $E = F \oplus F'$.

We shall now prove that these three conditions are equivalent.

Lemma 2.1. *Let $E = \sum_{i \in I}$ be a sum (not necessarily direct) of simple submodules. Then there exists a subset $J \subset I$ such that E is the direct sum $\bigoplus_{j \in J} E_j$.*

Proof. Let J be a maximal subset of I such that the sum $\sum_{j \in J} E_j$ is direct. We contend that this sum is in fact equal to E. It will suffice to prove that each E_i is contained in this sum. But the intersection of our sum with E_i is a submodule of E_i, hence equal to 0 or E_i. If it is equal to 0, then J is not maximal, since we can adjoin i to it. Hence E_i is contained in the sum, and our lemma is proved.

The lemma shows that **SS 1** implies **SS 2**. To see that **SS 2** implies **SS 3**, take a submodule F, and let J be a maximal subset of I such that the sum $F + \bigoplus_{j \in J} E_j$ is direct. The same reasoning as before shows that this sum is equal to E.

Finally assume *SS3*. To show **SS 1**, we shall first prove that every non-zero submodule of E contains a simple submodule. Let $v \in E$, $v \neq 0$. Then by definition, Rv is a principal submodule, and the kernel of the homomorphism

$$R \to Rv$$

is a left ideal $L \neq R$. Hence L is contained in a maximal left ideal $M \neq R$ (by Zorn's lemma). Then M/L is a maximal submodule of R/L (unequal to R/L), and hence Mv is a maximal submodule of Rv, unequal to Rv, corresponding to M/L under the isomorphism

$$R/L \to Rv.$$

We can write $E = Mv \oplus M'$ with some submodule M'. Then

$$Rv = Mv \oplus (M' \cap Rv),$$

because every element $x \in Rv$ can be written uniquely as a sum $x = \alpha v + x'$ with $\alpha \in M$ and $x' \in M'$, and $x' = x - \alpha v$ lies in Rv. Since Mv is maximal in Rv, it follows that $M' \cap Rv$ is simple, as desired.

Let E_0 be the submodule of E which is the sum of all simple submodules of E. If $E_0 \neq E$, then $E = E_0 \oplus F$ with $F \neq 0$, and there exists a simple submodule of F, contradicting the definition of E_0. This proves that **SS 3** implies **SS 1**.

A module E satisfying our three conditions is said to be **semisimple**.

Proposition 2.2. *Every submodule and every factor module of a semisimple module is semisimple.*

Proof. Let F be a submodule. Let F_0 be the sum of all simple submodules of F. Write $E = F_0 \oplus F'_0$. Every element x of F has a unique expression $x = x_0 + x'_0$ with $x_0 \in F_0$ and $x'_0 \in F'_0$. But $x'_0 = x - x_0 \in F$. Hence F is the direct sum

$$F = F_0 \oplus (F \cap F'_0).$$

We must therefore have $F_0 = F$, which is semisimple. As for the factor module, write $E = F \oplus F'$. Then F' is a sum of its simple submodules, and the canonical map $E \to E/F$ induces an isomorphism of F' onto E/F. Hence E/F is semisimple.

§3. THE DENSITY THEOREM

Let E be a semisimple R-module. Let $R' = R'(E)$ be the ring $\mathrm{End}_R(E)$. Then E is also a R'-module, the operation of R' on E being given by

$$(\varphi, x) \mapsto \varphi(x)$$

for $\varphi \in R'$ and $x \in E$. Each $\alpha \in R$ induces a R'-homomorphism $f_\alpha \colon E \to E$ by the map $f_\alpha(x) = \alpha x$. This is what is meant by the condition

$$\varphi(\alpha x) = \alpha\varphi(x).$$

We let $R'' = R''(E) = \mathrm{End}_{R'}(E)$. We call R' the **commutant** of R and R'' the **bicommutant**. Thus we get a ring-homomorphism

$$R \to \mathrm{End}_{R'}(E) = R''(E) = R''$$

by $\alpha \mapsto f_\alpha$. We now ask how big is the image of this ring-homomorphism. The density theorem states that it is quite big.

Lemma 3.1. *Let E be semisimple over R. Let $R' = \text{End}_R(E)$, $f \in \text{End}_{R'}(E)$ as above. Let $x \in R$. There exists an element $\alpha \in R$ such that $\alpha x = f(x)$.*

Proof. Since E is semisimple, we can write an R-direct sum

$$E = Rx \oplus F$$

with some submodule F. Let $\pi: E \to Rx$ be the projection. Then $\pi \in R'$, and hence

$$f(x) = f(\pi x) = \pi f(x).$$

This shows that $f(x) \in Rx$, as desired.

The density theorem generalizes the lemma by dealing with a finite number of elements of E instead of just one. For the proof, we use a diagonal trick.

Theorem 3.2. (Jacobson). *Let E be semisimple over R, and let $R' = \text{End}_R(E)$. Let $f \in \text{End}_{R'}(E)$. Let $x_1, \ldots, x_n \in E$. Then there exists an element $\alpha \in R$ such that*

$$\alpha x_i = f(x_i) \quad \textit{for} \quad i = 1, \ldots, n.$$

If E is finitely generated over R', then the natural map $R \to \text{End}_{R'}(E)$ is surjective.

Proof. For clarity of notation, we shall first carry out the proof in case E is simple. Let $f^{(n)} : E^{(n)} \to E^{(n)}$ be the product map, so that

$$f^{(n)}(y_1, \ldots, y_n) = (f(y_1), \ldots, f(y_n)).$$

Let $R'_n = \text{End}_R(E^{(n)})$. Then R'_n is none other than the ring of matrices with coefficients in R'. Since f commutes with elements of R' in its action on E, one sees immediately that $f^{(n)}$ is in $\text{End}_{R'_n}(E^{(n)})$. By the lemma, there exists an element $\alpha \in R$ such that

$$(\alpha x_1, \ldots, \alpha x_n) = (f(x_1), \ldots, f(x_n)),$$

which is what we wanted to prove.

When E is not simple, suppose that E is equal to a finite direct sum of simple submodules E_i (non-isomorphic), with multiplicities n_i:

$$E = E_1^{(n_1)} \oplus \cdots \oplus E_r^{(n_r)} \qquad (E_i \not\approx E_j \quad \text{if} \quad i \neq j),$$

then the matrices representing the ring of endomorphisms split according to blocks corresponding to the non-isomorphic simple components in our direct sum decomposition. Hence here again the argument goes through as before.

The main point is that $f^{(n)}$ lies in $\mathrm{End}_{R_n'}(E^{(n)})$, and that we can apply the lemma.

We add the observation that if E is finitely generated over R', then an element $f \in \mathrm{End}_{R'}(E)$ is determined by its value on a finite number of elements of E, so the asserted surjectivity $R \to \mathrm{End}_{R'}(E)$ follows at once. In the applications below, E will be a finite dimensional vector space over a field k, and R will be a k-algebra, so the finiteness condition is automatically satisfied.

The argument when E is an infinite direct sum would be similar, but the notation is disagreeable. However, in the applications we shall never need the theorem in any case other than the case when E itself is a finite direct sum of simple modules, and this is the reason why we first gave the proof in that case, and let the reader write out the formal details in the other cases, if desired.

Corollary 3.3. (Burnside's Theorem). *Let E be a finite-dimensional vector space over an algebraically closed field k, and let R be a subalgebra of $\mathrm{End}_k(E)$. If E is a simple R-module, then $R = \mathrm{End}_{R'}(E)$.*

Proof. We contend that $\mathrm{End}_R(E) = k$. At any rate, $\mathrm{End}_R(E)$ is a division ring R', containing k as a subring and every element of k commutes with every element of R'. Let $\alpha \in R'$. Then $k(\alpha)$ is a field. Furthermore, R' is contained in $\mathrm{End}_k(E)$ as a k-subspace, and is therefore finite dimensional over k. Hence $k(\alpha)$ is finite over k, and therefore equal to k since k is algebraically closed. This proves that $\mathrm{End}_R(E) = k$. Let now $\{v_1, \ldots, v_n\}$ be a basis of E over k. Let $A \in \mathrm{End}_k(E)$. According to the density theorem, there exists $\alpha \in R$ such that

$$\alpha v_i = A v_i \quad \text{for} \quad i = 1, \ldots, n.$$

Since the effect of A is determined by its effect on a basis, we conclude that $R = \mathrm{End}_k(E)$.

Corollary 3.3 is used in the following situation as in Exercise 8. Let E be a finite-dimensional vector space over field k. Let G be a submonoid of $GL(E)$ (multiplicative). A ***G*-invariant** subspace F of E is a subspace such that $\sigma F \subset F$ for all $\sigma \in G$. We say that E is ***G*-simple** if it has no G-invariant subspace other than 0 and E itself, and $E \neq 0$. Let $R = k[G]$ be the subalgebra of $\mathrm{End}_k(E)$ generated by G over k. Since we assumed that G is a monoid, it follows that R consists of linear combinations

$$\sum a_i \sigma_i$$

with $a_i \in k$ and $\sigma_i \in G$. Then we see that a subspace F of E is G-invariant if and only if it is R-invariant. Thus E is G-simple if and only if it is simple over R in the sense which we have been considering. We can then restate Burnside's theorem as he stated it:

Corollary 3.4. *Let E be a finite dimensional vector space over an algebraically closed field k, and let G be a (multiplicative) submonoid of $GL(E)$.*

If E is G-simple, then $k[G] = \text{End}_k(E)$.

When k is not algebraically closed, then we still get some result. Quite generally, let R be a ring and E a simple R-module. We have seen that $\text{End}_R(E)$ is a division ring, which we denote by D, and E is a vector space over D.

Let R be a ring, and E any R-module. We shall say that E is a **faithful** module if the following condition is satisfied. Given $\alpha \in R$ such that $\alpha x = 0$ for all $x \in E$, we have $\alpha = 0$. In the applications, E is a vector space over a field k, and we have a ring-homomorphism of R into $\text{End}_k(E)$. In this way, E is an R-module, and it is faithful if and only if this homomorphism is injective.

Corollary 3.5. **(Wedderburn's Theorem).** *Let R be a ring, and E a simple, faithful module over R. Let $D = \text{End}_R(E)$, and assume that E is finite dimensional over D. Then $R = \text{End}_D(E)$.*

Proof. Let $\{v_1, \ldots, v_n\}$ be a basis of E over D. Given $A \in \text{End}_D(E)$, by Theorem 3.2 there exists $\alpha \in R$ such that

$$\alpha v_i = A v_i \quad \text{for} \quad i = 1, \ldots, n.$$

Hence the map $R \to \text{End}_D(E)$ is surjective. Our assumption that E is faithful over R implies that it is injective, and our corollary is proved.

Example. Let R be a finite-dimensional algebra over a field k, and assume that R has a unit element, so is a ring. If R does not have any two-sided ideals other than 0 and R itself, then any nonzero module E over R is faithful, because the kernel of the homomorphism

$$R \to \text{End}_k(E)$$

is a two-sided ideal $\neq R$. If E is simple, then E is finite dimensional over k. Then D is a finite-dimensional division algebra over k. Wedderburn's theorem gives a representation of R as the ring of D-endomorphisms of E.

Under the assumption that R is finite dimensional, one can find a simple module simply by taking a minimal left ideal $\neq 0$. Such an ideal exists merely by taking a left ideal of minimal non-zero dimension over k. An even shorter proof of Wedderburn's theorem will be given below (Rieffel's theorem) in this case.

Corollary 3.6. *Let R be a ring, finite dimensional algebra over a field k which is algebraically closed. Let V be a finite dimensional vector space over k, with a simple faithful representation $\rho: R \to \text{End}_k(V)$. Then ρ is an isomorphism, in other words, $R \approx \text{Mat}_n(k)$.*

Proof. We apply Corollary 3.5, noting that D is finite dimensional over k. Given $\alpha \in D$, we note that $k(\alpha)$ is a commutative subfield of D, whence $k(\alpha) = k$ by assumption that k is algebraically closed, and the corollary follows.

Note. The corollary applies to simple rings, which will be defined below.

Suppose next that $V_1, \ldots, V_m$ are finite dimensional vector spaces over a field k, and that R is a k-algebra with representations

$$R \to \operatorname{End}_k(V_i), \ i = 1, \ldots, m,$$

so V_i is an R-module. If we let

$$E = V_1 \oplus \cdots \oplus V_m,$$

then E is finite over $R'(E)$, so we get the following consequence of Jacobson's density theorem.

Theorem 3.7. Existence of projection operators. *Let k be a field, R a k-algebra, and $V_1, \ldots, V_m$ finite dimensional k-spaces which are also simple R-modules, and such that V_i is not R-isomorphic to V_j for $i \neq j$. Then there exist elements $e_i \in R$ such that e_i acts as the identity on V_i and $e_i V_j = 0$ if $j \neq i$.*

Proof. We observe that the projection f_i from the direct sum E to the i-th factor is in $\operatorname{End}_{R'}(E)$, because if $\varphi \in R'$ then $\varphi(V_j) \subset V_j$ for all j. We may therefore apply the density theorem to conclude the proof.

Corollary 3.8. (Bourbaki). *Let k be a field of characteristic 0. Let R be a k-algebra, and let E, F be semisimple R-modules, finite dimensional over k. For each $\alpha \in R$, let α_E, α_F be the corresponding k-endomorphisms on E and F respectively. Suppose that the traces are equal; that is,*

$$\operatorname{tr}(\alpha_E) = \operatorname{tr}(\alpha_F) \textit{ for all } \alpha \in R.$$

Then E is isomorphic to F as R-module.

Proof. Each of E and F is isomorphic to a finite direct sum of simple R-modules, with certain multiplicities. Let V be a simple R-module, and suppose

$$E = V^{(n)} \oplus \text{direct summands not isomorphic to } V$$

$$F = V^{(m)} \oplus \text{direct summands not isomorphic to } V.$$

It will suffice to prove that $m = n$. Let e_V be the element of R found in Theorem 3.7 such that e_V acts as the identity on V, and is 0 on the other direct summands of E and F. Then

$$\operatorname{tr}(e_E) = n\dim_k(V) \quad \text{and} \quad \operatorname{tr}(e_F) = m\dim_k(V).$$

Since the traces are equal by assumption, it follows that $m = n$, thus concluding the proof. Note that the characteristic 0 is used here, because the values of the trace are in k.

Example. In the language of representations, suppose G is a monoid, and

we have two semisimple representations into finite dimensional k-spaces

$$\rho : G \to \mathrm{End}_k(E) \quad \text{and} \quad \rho' : G \to \mathrm{End}_k(F)$$

(so ρ and ρ' map G into the multiplicative monoid of End_k). Assume that $\mathrm{tr}\ \rho(\sigma) = \mathrm{tr}\ \rho'(\sigma)$ for all $\sigma \in G$. Then ρ and ρ' are isomorphic. Indeed, we let $R = k[G]$, so that ρ and ρ' extend to representations of R. By linearity, one has that $\mathrm{tr}\ \rho(\alpha) = \mathrm{tr}\ \rho'(\alpha)$ for all $\alpha \in R$, so one can apply Corollary 3.8.

§4. SEMISIMPLE RINGS

A ring R is called **semisimple** if $1 \neq 0$, and if R is semisimple as a left module over itself.

Proposition 4.1. *If R is semisimple, then every R-module is semisimple.*

Proof. An R-module is a factor module of a free module, and a free module is a direct sum of R with itself a certain number of times. We can apply Proposition 2.2 to conclude the proof.

Examples. 1) Let k be a field and let $R = \mathrm{Mat}_n(k)$ be the algebra of $n \times n$ matrices over k. Then R is semisimple, and actually simple, as we shall define and prove in §5, Theorem 5.5.

2) Let G be a finite group and suppose that the characteristic of k does not divide $\#(G)$. Then the group ring $k[G]$ is semisimple, as we shall prove in Chapter XVIII, Theorem 1.2.

3) The Clifford algebras C_n over the real numbers are semisimple. See Exercise 19 of Chapter XIX.

A left ideal of R is an R-module, and is thus called simple if it is simple as a module. Two ideals L, L' are called isomorphic if they are isomorphic as modules.

We shall now decompose R as a sum of its simple left ideals, and thereby get a structure theorem for R.

Let $\{L_i\}_{i \in I}$ be a family of simple left ideals, no two of which are isomorphic, and such that each simple left ideal is isomorphic to one of them. We say that this family is a family of representatives for the isomorphism classes of simple left ideals.

Lemma 4.2. *Let L be a simple left ideal, and let E be a simple R-module. If L is not isomorphic to E, then $LE = 0$.*

Proof. We have $RLE = LE$, and LE is a submodule of E, hence equal to

0 or E. Suppose $LE = E$. Let $y \in E$ be such that

$$Ly \neq 0.$$

Since Ly is a submodule of E, it follows that $Ly = E$. The map $\alpha \mapsto \alpha y$ of L into E is a homomorphism of L into E, which is surjective, and hence nonzero. Since L is simple, this homomorphism is an isomorphism.

Let

$$R_i = \sum_{L \approx L_i} L$$

be the sum of all simple left ideals isomorphic to L_i. From the lemma, we conclude that $R_i R_j = 0$ if $i \neq j$. This will be used constantly in what follows. We note that R_i is a left ideal, and that R is the sum

$$R = \sum_{i \in I} R_i,$$

because R is a sum of simple left ideals. Hence for any $j \in I$,

$$R_j \subset R_j R = R_j R_j \subset R_j,$$

the first inclusion because R contains a unit element, and the last because R_j is a left ideal. We conclude that R_j is also a right ideal, i.e. R_j is a two-sided ideal for all $j \in I$.

We can express the unit element 1 of R as a sum

$$1 = \sum_{i \in I} e_i$$

with $e_i \in R_i$. This sum is actually finite, almost all $e_i = 0$. Say $e_i \neq 0$ for indices $i = 1, \ldots, s$, so that we write

$$1 = e_1 + \cdots + e_s.$$

For any $x \in R$, write

$$x = \sum_{i \in I} x_i, \qquad x_i \in R_i.$$

For $j = 1, \ldots, s$ we have $e_j x = e_j x_j$ and also

$$x_j = 1 \cdot x_j = e_1 x_j + \cdots + e_s x_j = e_j x_j.$$

Furthermore, $x = e_1 x + \cdots + e_s x$. This proves that there is no index i other than $i = 1, \ldots, s$ and also that the i-th component x_i of x is uniquely determined as $e_i x = e_i x_i$. Hence the sum $R = R_1 + \cdots + R_s$ is direct, and furthermore, e_i is a unit element for R_i, which is therefore a ring. Since

$R_i R_j = 0$ for $i \neq j$, we find that in fact

$$R = \prod_{i=1}^{s} R_i$$

is a direct product of the rings R_i.

A ring R is said to be **simple** if it is semisimple, and if it has only one isomorphism class of simple left ideals. We see that we have proved a structure theorem for semisimple rings:

Theorem 4.3. *Let R be semisimple. Then there is only a finite number of non-isomorphic simple left ideals, say $L_1, \ldots, L_s$. If*

$$R_i = \sum_{L \approx L_i} L$$

is the sum of all simple left ideals isomorphic to L_i, then R_i is a two-sided ideal, which is also a ring (the operations being those induced by R), and R is ring isomorphic to the direct product

$$R = \prod_{i=1}^{s} R_i.$$

Each R_i is a simple ring. If e_i is its unit element, then $1 = e_1 + \cdots + e_s$, and $R_i = Re_i$. We have $e_i e_j = 0$ if $i \neq j$.

We shall now discuss modules.

Theorem 4.4. *Let R be semisimple, and let E be an R-module $\neq 0$. Then*

$$E = \bigoplus_{i=1}^{s} R_i E = \bigoplus_{i=1}^{s} e_i E,$$

and $R_i E$ is the submodule of E consisting of the sum of all simple submodules isomorphic to L_i.

Proof. Let E_i be the sum of all simple submodules of E isomorphic to L_i. If V is a simple submodule of E, then $RV = V$, and hence $L_i V = V$ for some i. By a previous lemma, we have $L_i \approx V$. Hence E is the direct sum of $E_1, \ldots, E_s$. It is then clear that $R_i E = E_i$.

Corollary 4.5. *Let R be semisimple. Every simple module is isomorphic to one of the simple left ideals L_i.*

Corollary 4.6. *A simple ring has exactly one simple module, up to isomorphism.*

Both these corollaries are immediate consequences of Theorems 4.3 and 4.4.

Proposition 4.7. *Let k be a field and E a finite dimensional vector space over k. Let S be a subset of* $\mathrm{End}_k(E)$. *Let R be the k-algebra generated by the elements of S. Then R is semisimple if and only if E is a semisimple R (or S) module.*

Proof. If R is semisimple, then E is semisimple by Proposition 4.1. Conversely, assume E semisimple as S-module. Then E is semisimple as R-module, and so is a direct sum

$$E = \bigoplus_{i=1}^{n} E_i$$

where each E_i is simple. Then for each i there exists an element $v_i \in E_i$ such that $E_i = Rv_i$. The map

$$x \mapsto (xv_1, \ldots, xv_n)$$

is a R-homomorphism of R into E, and is an injection since R is contained in $\mathrm{End}_k(E)$. Since a submodule of a semisimple module is semisimple by Proposition 2.2, the desired result follows.

§5. SIMPLE RINGS

Lemma 5.1. *Let R be a ring, and $\psi \in \mathrm{End}_R(R)$ a homomorphism of R into itself, viewed as R-module. Then there exists $\alpha \in R$ such that $\psi(x) = x\alpha$ for all $x \in R$.*

Proof. We have $\psi(x) = \psi(x \cdot 1) = x\psi(1)$. Let $\alpha = \psi(1)$.

Theorem 5.2. *Let R be a simple ring. Then R is a finite direct sum of simple left ideals. There are no two-sided ideals except 0 and R. If L, M are simple left ideals, then there exists $\alpha \in R$ such that $L\alpha = M$. We have $LR = R$.*

Proof. Since R is by definition also semisimple, it is a direct sum of simple left ideals, say $\bigoplus_{j \in J} L_j$. We can write 1 as a finite sum $1 = \sum_{j=1}^{m} \beta_j$, with $\beta_j \in L_j$. Then

$$R = \bigoplus_{j=1}^{m} R\beta_j = \bigoplus_{j=1}^{m} L_j.$$

This proves our first assertion. As to the second, it is a consequence of the third. Let therefore L be a simple left ideal. Then LR is a left ideal, because $RLR = LR$, hence (R being semisimple) is a direct sum of simple left ideals, say

$$LR = \bigoplus_{j=1}^{m} L_j, \qquad L = L_1.$$

Let M be a simple left ideal. We have a direct sum decomposition $R = L \oplus L'$. Let $\pi : R \to L$ be the projection. It is an R-endomorphism. Let $\sigma : L \to M$ be an isomorphism (it exists by Theorem 4.3). Then $\sigma \circ \pi : R \to R$ is an R-endomorphism. By the lemma, there exists $\alpha \in R$ such that

$$\sigma \circ \pi(x) = x\alpha \quad \text{for all} \quad x \in R.$$

Apply this to elements $x \in L$. We find

$$\sigma(x) = x\alpha \quad \text{for all} \quad x \in L.$$

The map $x \mapsto x\alpha$ is a R-homomorphism of L into M, is non-zero, hence is an isomorphism. From this it follows at once that $LR = R$, thereby proving our theorem.

Corollary 5.3. *Let R be a simple ring. Let E be a simple R-module, and L a simple left ideal of R. Then $LE = E$ and E is faithful.*

Proof. We have $LE = L(RE) = (LR)E = RE = E$. Suppose $\alpha E = 0$ for some $\alpha \in R$. Then $R\alpha RE = R\alpha E = 0$. But $R\alpha R$ is a two-sided ideal. Hence $R\alpha R = 0$, and $\alpha = 0$. This proves that E is faithful.

Theorem 5.4. (Rieffel). *Let R be a ring without two-sided ideals except* 0 *and R. Let L be a nonzero left ideal, $R' = \text{End}_R(L)$ and $R'' = \text{End}_{R'}(L)$. Then the natural map $\lambda : R \to R''$ is an isomorphism.*

Proof. The kernel of λ is a two-sided ideal, so λ is injective. Since LR is a two-sided ideal, we have $LR = R$ and $\lambda(L)\lambda(R) = \lambda(R)$. For any $x, y \in L$, and $f \in R''$, we have $f(xy) = f(x)y$, because right multiplication by y is an R-endomorphism of L. Hence $\lambda(L)$ is a left ideal of R'', so

$$R'' = R''\lambda(R) = R''\lambda(L)\lambda(R) = \lambda(L)\lambda(R) = \lambda(R),$$

as was to be shown.

In Rieffel's theorem, we do not need to assume that L is a simple module.

On the other hand, L is an ideal. So this theorem is not equivalent with previous ones of the same nature. In §7, we shall give a very general condition under which the canonical homomorphism

$$R \to R''$$

of a ring into the double endomorphism ring of a module is an isomorphism. This will cover all the previous cases.

As pointed out in the example following Wedderburn's theorem, Rieffel's theorem applies to give another proof when R is a finite-dimensional algebra (with unit) over a field k.

The next theorem gives a converse, showing that matrix rings over division algebras are simple.

Theorem 5.5. *Let D be a division ring, and E a finite-dimensional vector space over D. Let $R = \mathrm{End}_D(E)$. Then R is simple and E is a simple R-module. Furthermore, $D = \mathrm{End}_R(E)$.*

Proof. We first show that E is a simple R-module. Let $v \in E$, $v \neq 0$. Then v can be completed to a basis of E over D, and hence, given $w \in E$, there exists $\alpha \in R$ such that $\alpha v = w$. Hence E cannot have any invariant subspaces other than 0 or itself, and is simple over R. It is clear that E is faithful over R. Let $\{v_1, \dots, v_m\}$ be a basis of E over D. The map

$$\alpha \mapsto (\alpha v_1, \dots, \alpha v_m)$$

of R into $E^{(m)}$ is an R-homomorphism of R into $E^{(m)}$, and is injective. Given $(w_1, \dots, w_m) \in E^{(m)}$, there exists $\alpha \in R$ such that $\alpha v_i = w_i$ and hence R is R-isomorphic to $E^{(m)}$. This shows that R (as a module over itself) is isomorphic to a direct sum of simple modules and is therefore semisimple. Furthermore, all these simple modules are isomorphic to each other, and hence R is simple by Theorem 4.3.

There remains to prove that $D = \mathrm{End}_R(E)$. We note that E is a semisimple module over D since it is a vector space, and every subspace admits a complementary subspace. We can therefore apply the density theorem (the roles of R and D are now permuted!). Let $\varphi \in \mathrm{End}_R(E)$. Let $v \in E$, $v \neq 0$. By the density theorem, there exists an element $a \in D$ such that $\varphi(v) = av$. Let $w \in E$. There exists an element $f \in R$ such that $f(v) = w$. Then

$$\varphi(w) = \varphi(f(v)) = f(\varphi(v)) = f(av) = af(v) = aw.$$

Therefore $\varphi(w) = aw$ for all $w \in E$. This means that $\varphi \in D$, and concludes our proof.

Theorem 5.6. *Let k be a field and E a finite-dimensional vector space of*

dimension m over k. Let $R = \mathrm{End}_k(E)$. Then R is a k-space, and

$$\dim_k R = m^2.$$

Furthermore, m is the number of simple left ideals appearing in a direct sum decomposition of R as such a sum.

Proof. The k-space of k-endomorphisms of E is represented by the space of $m \times m$ matrices in k, so the dimension of R as a k-space is m^2. On the other hand, the proof of Theorem 5.5 showed that R is R-isomorphic as an R-module to the direct sum $E^{(m)}$. We know the uniqueness of the decomposition of a module into a direct sum of simple modules (Proposition 1.2), and this proves our assertion.

In the terminology introduced in §1, we see that the integer m in Theorem 5.6 is the length of R.

We can identify $R = \mathrm{End}_k(E)$ with the ring of matrices $\mathrm{Mat}_m(k)$, once a basis of E is selected. In that case, we can take the simple left ideals to be the ideals L_i ($i = 1, \ldots, m$) where a matrix in L_i has coefficients equal to 0 except in the i-th column. An element of L_1 thus looks like

$$\begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ a_{21} & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ a_{m1} & 0 & \cdots & 0 \end{pmatrix}$$

We see that R is the direct sum of the m columns.

We also observe that Theorem 5.5 implies the following:

If a matrix $M \in \mathrm{Mat}_m(k)$ commutes with all elements of $\mathrm{Mat}_m(k)$, then M is a scalar matrix.

Indeed, such a matrix M can then be viewed as an R-endomorphism of E, and we know by Theorem 5.5 that such an endomorphism lies in k. Of course, one can also verify this directly by a brute force computation.

§6. THE JACOBSON RADICAL, BASE CHANGE, AND TENSOR PRODUCTS

Let R be a ring and let M be a maximal left ideal. Then R/M is an R-module, and actually R/M is simple. Indeed, let $\bar{J}$ be a submodule of R/M with $\bar{J} \neq R/M$. Let J be its inverse image in R under the canonical homomorphism.

Then J is a left ideal $\neq M$ because $\bar{J} \neq R/M$, so $J = R$ and $\bar{J} = 0$. Conversely, let E be a simple R-module and let $v \in E$, $v \neq 0$. Then Rv is a submodule $\neq 0$ of E, and hence $Rv = E$. Let M be the kernel of the homomorphism $x \mapsto xv$. Then M is a left ideal, and M is maximal; otherwise there is a left ideal M' with $R \supset M' \supset M$ and $M' \neq R, \neq M$. Then $R/M \approx E$ and R/M' is a non-zero homomorphic image of E, which cannot exist since E is simple (Schur's lemma, Proposition 1.1). Thus we obtain a bijection between maximal left ideals and simple R-modules (up to isomorphism).

We define the **Jacobson radical** of R to be the left ideal N which is the intersection of all maximal left ideals of R. We may also denote $N = \text{Rad}(R)$.

Theorem 6.1. **(a)** *For every simple R-module we have* $NE = 0$.

(b) *The radical N is a two-sided ideal, containing all nilpotent two-sided ideals.*

(c) *Let R be a finite dimensional algebra over field k. Its radical is* $\{0\}$, *if and only if R is semisimple.*

(d) *If R is a finite dimensional algebra over a field k, then its radical N is nilpotent* (*i.e.* $N^r = 0$ *for some positive integer r*).

These statements are easy to prove, and hints will be given appropriately. See Exercises 1 through 5.

Observe that under finite dimensionality conditions, the radical's being 0 gives us a useful criterion for a ring to be semisimple, which we shall use in the next result.

Theorem 6.2. *Let A be a semisimple algebra, finite dimensional over a field k. Let K be a finite separable extension of k. Then* $K \otimes_k A$ *is a semisimple over K.*

Proof. In light of the radical criterion for semisimplicity, it suffices to prove that $K \otimes_k A$ has zero radical, and it suffices to do so for an even larger extension than K, so that we may assume K is Galois over k, say with Galois group G. Then G operates on $K \otimes A$ by

$$\sigma(x \otimes a) = \sigma x \otimes a \quad \text{for} \quad x \in K \quad \text{and} \quad a \in A.$$

Let N be the radical of $K \otimes A$. Since N is nilpotent, it follows that σN is also nilpotent for all $\sigma \in G$, whence $\sigma N = N$ because N is the maximal nilpotent ideal (Exercise 5). Let $\{\alpha_1, \ldots, \alpha_m\}$ be a basis of A over k. Suppose N contains the element

$$\xi = \sum x_i \otimes \alpha_i \neq 0 \quad \text{with} \quad x_i \in K.$$

For every $y \in K$ the element $(y \otimes 1)\xi = \sum yx_i \otimes \alpha_i$ also lies in N. Then

$$\text{trace}((y \otimes 1)\xi) = \sum \sigma\xi = \sum \text{Tr}(yx_i) \otimes \alpha_i = \sum 1 \otimes \alpha_i \text{Tr}(yx_i)$$

also lies in N, and lies in $1 \otimes A \approx A$, thus proving the theorem.

Remark. For the case when A is a finite extension of k, compare with Exercises 1, 2, 3 of Chapter XVI.

Let A be a semisimple algebra, finite dimensional over a field k. Then by Theorem 6.2 the extension of scalars $A \otimes_k k^a$ is semisimple if k is perfect. In general, an algebra A over k is said to be **absolutely semisimple** if $A \otimes_k k^a$ is semisimple.

We now look at semisimple algebras over an algebraically closed field.

Theorem 6.3. *Let A, B be simple algebras, finite dimensional over a field k which is algebraically closed. Then $A \otimes_k B$ is also simple. We have $A \approx \text{End}_k(V)$ and $B \approx \text{End}_k(W)$ where V, W are finite dimensional vector spaces over k, and there is a natural isomorphism*

$$A \otimes_k B \approx \text{End}_k(V \otimes_k W) \approx \text{End}_k(V) \otimes_k \text{End}_k(W).$$

Proof. The formula is a special case of Theorem 2.5 of Chapter XVI, and the isomorphisms $A \approx \text{End}_k(V)$, $B \approx \text{End}_k(W)$ exist by Wedderburn's theorem or its corollaries.

Let A be an algebra over k and let F be an extension field of k. We denote by A_F the extension of scalars

$$A_F = A \otimes_k F.$$

Thus A_F is an algebra over F. As an exercise, prove that if k is the center of A, then F is the center of A_F. (Here we identify F with $1 \otimes F$.)

Let A, B be algebras over k. We leave to the reader the proof that for every extension field F of k, we have a natural isomorphism

$$(A \otimes_k B)_F = A_F \otimes_F B_F.$$

We apply the above considerations to the tensor product of semisimple algebras.

Theorem 6.4. *Let A, B be absolutely semisimple algebras finite dimensional over a field k. Then $A \otimes_k B$ is absolutely semisimple.*

Proof. Let $F = k^a$. Then A_F is semisimple by hypothesis, so it is a direct product of simple algebras, which are matrix algebras, and in particular we can apply Theorem 6.3 to see that $A_F \otimes_F B_F$ has no radical. Hence $A \otimes_k B$ has no radical (because if N is its radical, then $N \otimes_k F = N_F$ is a nilpotent ideal of $A_F \otimes_F B_F$), whence $A \otimes_k B$ is semisimple by Theorem 6.1(c).

Remark. We have proved the above tensor product theorems rapidly in special cases, which are already important in various applications. For a more general treatment, I recommend Bourbaki's *Algebra*, Chapter VIII, which gives an exhaustive treatment of tensor products of semisimple and simple algebras.

§7. BALANCED MODULES

Let R be a ring and E a module. We let $R'(E) = \mathrm{End}_R(E)$ and

$$R''(E) = \mathrm{End}_{R'}(E).$$

Let $\lambda : R \to R''$ be the natural homomorphism such that $\lambda_x(v) = xv$ for $x \in R$ and $v \in E$. If λ is an isomorphism, we shall say that E is **balanced**. We shall say that E is a **generator** (for R-modules) if every module is a homomorphic image of a (possibly infinite) direct sum of E with itself. For example, R is a generator.

More interestingly, in Rieffel's Theorem 5.4, the left ideal L is a generator, because $LR = R$ implies that there is a surjective homomorphism $L \times \cdots \times L \to R$ since we can write 1 as a finite combination

$$1 = x_1a_1 + \cdots + x_na_n \text{ with } x_i \in L \text{ and } a_i \in R.$$

The map $(x_1, \ldots, x_n) \mapsto x_1a_1 + \cdots + x_na_n$ is a R-homomorphism of left module onto R.

If E is a generator, then there is a surjective homomorphism $E^{(n)} \to R$ (we can take n finite since R is finitely generated, by one element 1).

Theorem 7.1. **(Morita).** *Let E be an R-module. Then E is a generator if and only if E is balanced and finitely generated projective over $R'(E)$.*

Proof. We shall prove half of the theorem, leaving the other half to the reader, using similar ideas (see Exercise 12). So we assume that E is a generator, and we prove that it satisfies the other properties by arguments due to Faith.

We first prove that for any module F, $R \oplus F$ is balanced. We identify R and F as the submodules $R \oplus 0$ and $0 \oplus F$ of $R \oplus F$, respectively. For $w \in F$, let $\psi_w : R \oplus F \to F$ be the map $\psi_w(x + v) = xw$. Then any $f \in R''(R \oplus F)$ commutes with π_1, π_2, and each ψ_w. From this we see at once that $f(x + v) = f(1)(x + v)$ and hence that $R \oplus F$ is balanced. Let E be a generator, and $E^{(n)} \to R$ a surjective homomorphism. Since R is free, we can write $E^{(n)} \approx R \oplus F$ for some module F, so that $E^{(n)}$ is balanced, Let $g \in R'(E)$. Then $g^{(n)}$ commutes with every element $\varphi = (\varphi_{ij})$ in $R'(E^{(n)})$ (with components $\varphi_{ij} \in R'(E)$), and hence there is some $x \in R$ such that $g^{(n)} = \lambda_x^{(n)}$. Hence $g = \lambda_x$, thereby proving that E is balanced, since λ is obviously injective.

To prove that E is finitely generated over $R'(E)$, we have

$$R'(E)^{(n)} \approx \mathrm{Hom}_R(E^{(n)}, E) \approx \mathrm{Hom}_R(R, E) \oplus \mathrm{Hom}_R(F, E)$$

as additive groups. This relation also obviously holds as R'-modules if we define the operation of R' to be composition of mappings (on the left). Since $\mathrm{Hom}_R(R, E)$ is R'-isomorphic to E under the map $h \mapsto h(1)$, it follows that E is an R'-homomorphic image of $R'^{(n)}$, whence finitely generated over R'. We also see that E is a direct summand of the free R'-module $R'^{(n)}$ and is therefore projective over $R'(E)$. This concludes the proof.

EXERCISES

The radical

1. (a) Let R be a ring. We define the **radical** of R to be the left ideal N which is the intersection of all maximal left ideals of R. Show that $NE = 0$ for every simple R-module E. Show that N is a two-sided ideal. (b) Show that the radical of R/N is 0.

2. A ring is said to be **Artinian** if every descending sequence of left ideals $J_1 \supset J_2 \supset \cdots$ with $J_i \neq J_{i+1}$ is finite. (a) Show that a finite dimensional algebra over a field is Artinian. (b) If R is Artinian, show that every non-zero left ideal contains a simple left ideal. (c) If R is Artinian, show that every non-empty set of ideals contains a minimal ideal.

3. Let R be Artinian. Show that its radical is 0 if and only if R is semisimple. [*Hint*: Get an injection of R into a direct sum $\bigoplus R/M_i$ where $\{M_i\}$ is a finite set of maximal left ideals.]

4. **Nakayama's lemma.** Let R be any ring and M a finitely generated module. Let N be the radical of R. If $NM = M$ show that $M = 0$. [*Hint*: Observe that the proof of Nakayama's lemma still holds.]

5. (a) Let J be a two-sided nilpotent ideal of R. Show that J is contained in the radical. (b) Conversely, assume that R is Artinian. Show that its radical is nilpotent, i.e., that there exists an integer $r \geqq 1$ such that $N^r = 0$. [*Hint*: Consider the descending sequence of powers N^r, and apply Nakayama to a minimal finitely generated left ideal $L \subset N^\infty$ such that $N^\infty L \neq 0$.

6. Let R be a semisimple commutative ring. Show that R is a direct product of fields.

7. Let R be a finite dimensional commutative algebra over a field k. If R has no nilpotent element $\neq 0$, show that R is semisimple.

8. (Kolchin) Let E be a finite-dimensional vector space over a field k. Let G be a subgroup of $GL(E)$ such that every element $A \in G$ is of type $I + N$ where N is nilpotent. Assume $E \neq 0$. Show that there exists an element $v \in E$, $v \neq 0$ such that $Av = v$ for all $A \in G$. [*Hint*: First reduce the question to the case when k is algebraically closed by showing that the problem amounts to solving linear equations. Secondly, reduce it to the case when E is a simple $k[G]$-module. Combining Burnside's theorem with the fact that $\mathrm{tr}(A) = \mathrm{tr}(I)$ for all $A \in G$, show that if $A_0 \in G$, $A_0 = I + N$, then $\mathrm{tr}(NX) = 0$ for all $X \in \mathrm{End}_k(E)$, and hence that $N = 0$, $A_0 = I$.]

Semisimple operations

9. Let E be a finite dimensional vector space over a field k. Let R be a semisimple subalgebra of $\mathrm{End}_k(E)$. Let $a, b \in R$. Assume that

$$\mathrm{Ker}\, b_E \supset \mathrm{Ker}\, a_E,$$

where b_E is multiplication by b on E and similarly for a_E. Show that there exists an element $s \in R$ such that $sa = b$. [*Hint*: Reduce to R simple. Then $R = \mathrm{End}_D(E_0)$ and $E = E_0^{(n)}$. Let $v_1, \ldots, v_r \in E$ be a D-basis for aE. Define s by $s(av_i) = bv_i$ and

extend s by D-linearity. Then $sa_E = b_E$, so $sa = b$.]

10. Let E be a finite-dimensional vector space over a field k. Let $A \in \mathrm{End}_k(E)$. We say that A is **semisimple** if E is a semisimple A-space, or equivalently, let R be the k-algebra generated by A, then E is semisimple over R. Show that A is semisimple if and only if its minimal polynomial has no factors of multiplicity > 1 over k.

11. Let E be a finite-dimensional vector space over a field k, and let S be a commutative set of endomorphisms of E. Let $R = k[S]$. Assume that R is semisimple. Show that every subset of S is semisimple.

12. Prove that an R-module E is a generator if and only if it is balanced, and finitely generated projective over $R'(E)$. Show that Theorem 5.4 is a consequence of Theorem 7.1.

13. Let A be a principal ring with quotient field K. Let A^n be n-space over A, and let

$$T = A^n \oplus A^n \oplus \cdots \oplus A^n$$

be the direct sum of A^n with itself r times. Then T is free of rank nr over A. If we view elements of A^n as column vectors, then T is the space of $n \times r$ matrices over A. Let $M = \mathrm{Mat}_n(A)$ be the ring of $n \times n$ matrices over A, operating on the left of T. By a **lattice** L in T we mean an A-submodule of rank nr over A. Prove that any such lattice which is M-stable is M-isomorphic to T itself. Thus there is just one M-isomorphism class of lattices. [*Hint*: Let $g \in M$ be the matrix with 1 in the upper left corner and 0 everywhere else, so g is a projection of A^n on a 1-dimensional subspace. Then multiplication on the left g: $T \to A_r$ maps T on the space of $n \times r$ matrices with arbitrary first row and 0 everywhere else. Furthermore, for any lattice L in T the image gL is a lattice in A_r, that is a free A-submodule of rank r. By elementary divisors there exists an $r \times r$ matrix Q such that

$$gL = A_r Q \qquad \text{(multiplication on the right).}$$

Then show that $TQ = L$ and that multiplication by Q on the right is an M-isomorphism of T with L.]

14. Let F be a field. Let $\mathfrak{n} = \mathfrak{n}(F)$ be the vector space of strictly upper triangular $n \times n$ matrices over F. Show that $\mathfrak{n}$ is actually an algebra, and all elements of $\mathfrak{n}$ are nilpotent (some positive integral power is 0).

15. **Conjugation representation.** Let A be the multiplicative group of diagonal matrices in F with non-zero diagonal components. For $a \in A$, the **conjugation action** of a on $\mathrm{Mat}_n(F)$ is denoted by $\mathbf{c}(a)$, so $\mathbf{c}(a)M = aMa^{-1}$ for $M \in \mathrm{Mat}_n(F)$. (a) Show that $\mathfrak{n}$ is stable under this action. (b) Show that $\mathfrak{n}$ is semisimple under this action. More precisely, for $1 \leqq i < j \leqq n$, let E_{ij} be the matrix with (ij)-component 1, and all other components 0. Then these matrices E_{ij} form a basis for $\mathfrak{n}$ over F, and each E_{ij} is an eigenvector for the conjugation action, namely for $a = \mathrm{diag}(a_1, \dots, a_n)$, we have

$$aE_{ij}a^{-1} = (a_i/a_j)E_{ij},$$

so the corresponding character χ_{ij} is given by $\chi_{ij}(a) = a_i/a_j$. (c) Show that $\mathrm{Mat}_n(F)$ is semisimple, and in fact is equal to $\mathfrak{d} \oplus \mathfrak{n} \oplus {}^t\mathfrak{n}$, where $\mathfrak{d}$ is the space of diagonal matrices.

CHAPTER XVIII

Representations of Finite Groups

The theory of group representations occurs in many contexts. First, it is developed for its own sake: determine all irreducible representations of a given group. See for instance Curtis-Reiner's *Methods of Representation Theory* (Wiley-Interscience, 1981). It is also used in classifying finite simple groups. But already in this book we have seen applications of representations to Galois theory and the determination of the Galois group over the rationals. In addition, there is an analogous theory for topological groups. In this case, the closest analogy is with compact groups, and the reader will find a self-contained treatment of the compact case entirely similar to §5 of this chapter in my book $\mathbf{SL}_2(\mathbf{R})$ (Springer Verlag), Chapter II, §2. Essentially, finite sums are replaced by integrals, otherwise the formalism is the same. The analysis comes only in two places. One of them is to show that every irreducible representation of a compact group is finite dimensional; the other is Schur's lemma. The details of these extra considerations are carried out completely in the above-mentioned reference. I was careful to write up §5 with the analogy in mind.

Similarly, readers will find analogous material on induced representations in $\mathbf{SL}_2(\mathbf{R})$, Chapter III, §2 (which is also self-contained).

Examples of the general theory come in various shapes. Theorem 8.4 may be viewed as an example, showing how a certain representation can be expressed as a direct sum of induced representations from 1-dimensional representations. Examples of representations of S_3 and S_4 are given in the exercises. The entire last section works out completely the simple characters for the group $GL_2(\mathbf{F})$ when $\mathbf{F}$ is a finite field, and shows how these characters essentially come from induced characters.

For other examples also leading into Lie groups, see W. Fulton and J. Harris, *Representation Theory*, Springer Verlag 1991.

§1. REPRESENTATIONS AND SEMISIMPLICITY

Let R be a commutative ring and G a group. We form the group algebra $R[G]$. As explained in Chapter II, §3 it consists of all formal linear combinations

$$\sum_{\sigma \in G} a_\sigma \sigma$$

with coefficients $a_\sigma \in R$, almost all of which are 0. The product is taken in the natural way,

$$\left(\sum_{\sigma \in G} a_\sigma \sigma\right)\left(\sum_{\tau \in G} b_\tau \tau\right) = \sum_{\sigma, \tau} a_\sigma b_\tau \sigma\tau.$$

Let E be an R-module. Every algebra-homomorphism

$$R[G] \to \operatorname{End}_R(E)$$

induces a group-homomorphism

$$G \to \operatorname{Aut}_R(E),$$

and thus a representation of the ring $R[G]$ in E gives rise to a representation of the group. Given such representations, we also say that $R[G]$, or G, **operate** on E. We note that the representation makes E into a module over the ring $R[G]$.

Conversely, given a representation of the group, say $\rho : G \to \operatorname{Aut}_R(E)$, we can extend ρ to a representation of $R[G]$ as follows. Let $\alpha = \sum a_\sigma \sigma$ and $x \in E$. We define

$$\rho(\alpha)x = \sum a_\sigma \rho(\sigma)x.$$

It is immediately verified that ρ has been extended to a ring-homomorphism of $R[G]$ into $\operatorname{End}_R(E)$. We say that ρ is **faithful** on G if the map $\rho : G \to \operatorname{Aut}_R(E)$ is injective. The extension of ρ to $R[G]$ may not be faithful, however.

Given a representation of G on E, we often write simply σx instead of $\rho(\sigma)x$, whenever we deal with a fixed representation throughout a discussion.

An R-module E, together with a representation ρ, will be called a **G-module**, or **G-space**, or also a (G, R)-module if we wish to specify the ring R. If E, F are G-modules, we recall that a G-homomorphism $f : E \to F$ is an R-linear map such that $f(\sigma x) = \sigma f(x)$ for all $x \in E$ and $\sigma \in G$.

Given a G-homomorphism $f : E \to F$, we note that the kernel of f is a G-submodule of E, and that the R-factor module $F/f(E)$ admits an operation of G in a unique way such that the canonical map $F \to F/f(E)$ is a G-homomorphism.

By a **trivial** representation $\rho : G \to \operatorname{Aut}_R(E)$, we shall mean the representation such that $\rho(G) = 1$. A representation is trivial if and only if $\sigma x = x$ for all $x \in E$. We also say in that case that G **operates trivially**.

We make R into a G-module by making G act trivially on R.

We shall now discuss systematically the representations which arise from a given one, on Hom, the dual, and the tensor product. This pattern will be repeated later when we deal with induced representations.

First, $\mathrm{Hom}_R(E, F)$ is a G-module under the action defined for $f \in \mathrm{Hom}_R(E, F)$ by

$$([\sigma]f)(x) = \sigma f(\sigma^{-1}x).$$

The conditions for an operation are trivially verified. Note the σ^{-1} inside the expression. We shall usually omit parentheses, and write simply $[\sigma]f(x)$ for the left-hand side. We note that f is a G-homomorphism if and only if $[\sigma]f = f$ for all $\sigma \in G$.

We are particularly concerned when $F = R$ (so with trivial action), in which case $\mathrm{Hom}_R(E, R) = E^\vee$ is the dual module. In the terminology of representations, if $\rho \colon G \to \mathrm{Aut}_R(E)$ is a representation of G on E, then the action we have just described gives a representation denoted by

$$\rho^\vee : G \to \mathrm{Aut}_R(E^\vee),$$

and called the **dual representation** (also called contragredient (ugh!) in the literature).

Suppose now that the modules E, F are free and finite dimensional over R. Let ρ be representation of G on E. Let M be the matrix of $\rho(\sigma)$ with respect to a basis, and let $M^\vee$ be the matrix of $\rho^\vee(\sigma)$ with respect to the dual basis. Then it is immediately verified that

$$M^\vee = {}^tM^{-1}. \tag{1}$$

Next we consider the tensor product instead of Hom. Let E, E' be (G, R)-modules. We can form their tensor product $E \otimes E'$, always taken over R. Then there is a unique action of G on $E \otimes E'$ such that for $\sigma \in G$ we have

$$\sigma(x \otimes x') = \sigma x \otimes \sigma x'.$$

Suppose that E, F are finite free over R. Then the R-isomorphism

$$E^\vee \otimes F \approx \mathrm{Hom}_R(E, F) \tag{2}$$

of Chapter XVI, Corollary 5.5, is immediately verified to be a G-isomorphism.

Whether E is free or not, we define the G-**invariant** submodule of E to be $\mathrm{inv}_G(E) = R$-submodule of elements $x \in E$ such that $\sigma x = x$ for all $\sigma \in G$. If E, F are free then we have an R-isomorphism

$$\mathrm{inv}_G(E^\vee \otimes F) \approx \mathrm{Hom}_G(E, F). \tag{3}$$

If $\rho: G \to \mathrm{Aut}_R(E)$ and $\rho': G \to \mathrm{Aut}_R(E')$ are representations of G on E and E' respectively, then we define their **sum** $\rho \oplus \rho'$ to be the representation on the direct sum $E \oplus E'$, with $\sigma \in G$ acting componentwise. Observe that G-isomorphism classes of representations have an additive monoid structure under this direct sum, and also have an associative multiplicative structure under the tensor product. With the notation of representations, we denote this product by $\rho \otimes \rho'$. This product is distributive with respect to the addition (direct sum).

If G is a finite group, and E is a G-module, then we can define the **trace** $\mathrm{Tr}_G: E \to E$ which is an R-homomorphism, namely

$$\mathrm{Tr}_G(x) = \sum_{\sigma \in G} \sigma x.$$

We observe that $\mathrm{Tr}_G(x)$ lies in $\mathrm{inv}_G(E)$, i.e. is fixed under the operation of all elements of G. This is because

$$\tau\, \mathrm{Tr}_G(x) = \sum_{\sigma \in G} \tau\sigma x,$$

and multiplying by τ on the left permutes the elements of G.

In particular, if $f: E \to F$ is an R-homomorphism of G-modules, then $\mathrm{Tr}_G(f): E \to F$ is a G-homomorphism.

Proposition 1.1. *Let G be a finite group and let E', E, F, F' be G-modules. Let*

$$E' \xrightarrow{\varphi} E \xrightarrow{f} F \xrightarrow{\psi} F'$$

be R-homomorphisms, and assume that φ, ψ are G-homomorphisms. Then

$$\mathrm{Tr}_G(\psi \circ f \circ \varphi) = \psi \circ \mathrm{Tr}_G(f) \circ \varphi.$$

Proof. We have

$$\mathrm{Tr}_G(\psi \circ f \circ \varphi) = \sum_{\sigma \in G} \sigma(\psi \circ f \circ \varphi) = \sum_{\sigma \in G} (\sigma\psi) \circ (\sigma f) \circ (\sigma\varphi)$$

$$= \psi \circ \left(\sum_{\sigma \in G} \sigma f \right) \circ \varphi = \psi \circ \mathrm{Tr}_G(f) \circ \varphi.$$

Theorem 1.2. (Maschke). *Let G be a finite group of order n, and let k be a field whose characteristic does not divide n. Then the group ring $k[G]$ is semisimple.*

Proof. Let E be a G-module, and F a G-submodule. Since k is a field, there exists a k-subspace F' such that E is the k-direct sum of F and F'. We let the k-linear map $\pi: E \to F$ be the projection on F. Then $\pi(x) = x$ for all $x \in F$.

Let

$$\varphi = \frac{1}{n}\operatorname{Tr}_G(\pi).$$

We have then two G-homomorphisms

$$0 \to F \underset{\varphi}{\overset{j}{\rightleftarrows}} E$$

such that j is the inclusion, and $\varphi \circ j = \mathrm{id}$. It follows that E is the G-direct sum of F and Ker φ, thereby proving that $k[G]$ is semisimple.

Except in §7 we denote by G a finite group, and we denote E, F finite dimensional k-spaces, where k is a field of characteristic not dividing $\#(G)$. We usually denote $\#(G)$ by n.

§2. CHARACTERS

Let $\rho: k[G] \to \mathrm{End}_k(E)$ be a representation. By the **character** χ_ρ of the representation, we shall mean the k-valued function

$$\chi_\rho: k[G] \to k$$

such that $\chi_\rho(\alpha) = \operatorname{tr} \rho(\alpha)$ for all $\alpha \in k[G]$. The trace here is the trace of an endomorphism, as defined in Chapter XIII, §3. If we select a basis for E over k, it is the trace of the matrix representing $\rho(\alpha)$, i.e., the sum of the diagonal elements. We have seen previously that the trace does not depend on the choice of the basis. We sometimes write χ_E instead of χ_ρ.

We also call E the **representation space** of ρ.

By the **trivial character** we shall mean the character of the representation of G on the k-space equal to k itself, such that $\sigma x = x$ for all $x \in k$. It is the function taking the value 1 on all elements of G. We denote it by χ_0 or also by 1_G if we need to specify the dependence on G.

We observe that characters are functions on G, and that the values of a character on elements of $k[G]$ are determined by its values on G (the extension from G to $k[G]$ being by k-linearity).

We say that two representations ρ, φ of G on spaces E, F are **isomorphic** if there is a G-isomorphism between E and F. We then see that if ρ, φ are isomorphic representations, then their characters are equal. (Put in another way, if E, F are G-spaces and are G-isomorphic, then $\chi_E = \chi_F$.) In everything that follows, we are interested only in isomorphism classes of representations.

If E, F are G-spaces, then their direct sum $E \oplus F$ is also a G-space, the operation of G being componentwise. If $x \oplus y \in E \oplus F$ with $x \in E$ and $y \in F$, then $\sigma(x \oplus y) = \sigma x \oplus \sigma y$.

Similarly, the tensor product $E \otimes_k F = E \otimes F$ is a G-space, the operation of G being given by $\sigma(x \otimes y) = \sigma x \otimes \sigma y$.

Proposition 2.1. *If E, F are G-spaces, then*

$$\chi_E + \chi_F = \chi_{E \oplus F} \quad \textit{and} \quad \chi_E \chi_F = \chi_{E \otimes F}.$$

If $\chi^\vee$ denotes the character of the dual representation on $E^\vee$, then

$$\begin{aligned} \chi^\vee(\sigma) &= \chi(\sigma^{-1}) \\ &= \overline{\chi(\sigma)} \text{ if } k = \mathbf{C}. \end{aligned}$$

Proof. The first relation holds because the matrix of an element σ in the representation $E \oplus F$ decomposes into blocks corresponding to the representation in E and the representation in F. As to the second, if $\{v_i\}$ is a basis of E and $\{w_j\}$ is a basis of F over k, then we know that $\{v_i \otimes w_j\}$ is a basis of $E \otimes F$. Let $(a_{i\nu})$ be the matrix of σ with respect to our basis of E, and $(b_{j\mu})$ its matrix with respect to our basis of F. Then

$$\begin{aligned} \sigma(v_i \otimes w_j) = \sigma v_i \otimes \sigma w_j &= \sum_\nu a_{i\nu} v_\nu \otimes \sum_\mu b_{j\mu} w_\mu \\ &= \sum_{\nu, \mu} a_{i\nu} b_{j\mu} v_\nu \otimes w_\mu. \end{aligned}$$

By definition, we find

$$\chi_{E \otimes F}(\sigma) = \sum_i \sum_j a_{ii} b_{jj} = \chi_E(\sigma)\chi_F(\sigma),$$

thereby proving the statement about tensor products. The statement for the character of the dual representation follows from the formula for the matrix ${}^tM^{-1}$ given in §1. The value given as the complex conjugate in case $k = \mathbf{C}$ will be proved later in Corollary 3.2.

So far, we have defined the notion of character associated with a representation. It is now natural to form linear combinations of such characters with more general coefficients than positive integers. Thus by a **character** of G we shall mean a function on G which can be written as a linear combination of characters of representations with arbitrary integer coefficients. The characters associated with representations will be called **effective characters**. Everything we have defined of course depends on the field k, and we shall add **over k** to our expressions if we need to specify the field k.

We observe that the characters form a ring in view of Proposition 2.1. For most of our work we do not need the multiplicative structure, only the additive one.

By a **simple** or **irreducible character** of G one means the character of a simple representation (i.e., the character associated with a simple $k[G]$-module).

Taking into account Theorem 1.2, and the results of the preceding chapter concerning the structure of simple and semisimple modules over a semisimple ring (Chapter XVII, §4) we obtain:

Theorem 2.2. *There are only a finite number of simple characters of G (over k). The characters of representations of G are the linear combinations of the simple characters with integer coefficients $\geqq 0$.*

We shall use the direct product decomposition of a semisimple ring. We have

$$k[G] = \prod_{i=1}^{s} R_i$$

where each R_i is simple, and we have a corresponding decomposition of the unit element of $k[G]$:

$$1 = e_1 + \cdots + e_s,$$

where e_i is the unit element of R_i, and $e_i e_j = 0$ if $i \neq j$. Also, $R_i R_j = 0$ if $i \neq j$. *We note that $s = s(k)$ depends on k.*

If L_i denotes a typical simple module for R_i (say one of the simple left ideals), we let χ_i be the character of the representation on L_i.

We observe that $\chi_i(\alpha) = 0$ for all $\alpha \in R_j$ if $i \neq j$. This is a fundamental relation of orthogonality, which is obvious, but from which all our other relations will follow.

Theorem 2.3. *Assume that k has characteristic 0. Then every effective character has a unique expression as a linear combination*

$$\chi = \sum_{i=1}^{s} n_i \chi_i, \qquad n_i \in \mathbf{Z},\ n_i \geqq 0,$$

where $\chi_1, \ldots, \chi_s$ are the simple characters of G over k. Two representations are isomorphic if and only if their associated characters are equal.

Proof. Let E be the representation space of χ. Then by Theorem 4.4 of Chapter XVII,

$$E \approx \bigoplus_{i=1}^{s} n_i L_i.$$

The sum is finite because we assume throughout that E is finite dimensional. Since e_i acts as a unit element on L_i, we find

$$\chi_i(e_i) = \dim_k L_i.$$

We have already seen that $\chi_i(e_j) = 0$ if $i \neq j$. Hence

$$\chi(e_i) = n_i \dim_k L_i.$$

Since $\dim_k L_i$ depends only on the structure of the group algebra, we have recovered the multiplicities $n_1, \ldots, n_s$. Namely, n_i is the number of times that L_i occurs (up to an isomorphism) in the representation space of χ, and is the value of $\chi(e_i)$ divided by $\dim_k L_i$ (we are in characteristic 0). This proves our theorem.

As a matter of definition, in Theorem 2.3 we call n_i the **multiplicity** of χ_i in χ. In both corollaries, we continue to assume that k has characteristic 0.

Corollary 2.4. *As functions of G into k, the simple characters*

$$\chi_1, \ldots, \chi_s$$

are linearly independent over k.

Proof. Suppose that $\sum a_i \chi_i = 0$ with $a_i \in k$. We apply this expression to e_j and get

$$0 = (\sum a_i \chi_i)(e_j) = a_j \dim_k L_j.$$

Hence $a_j = 0$ for all j.

In characteristic 0 we define the **dimension** of an effective character to be the dimension of the associated representation space.

Corollary 2.5. *The function* dim *is a homomorphism of the monoid of effective characters into* $\mathbf{Z}$.

Example. Let G be a cyclic group of order equal to a prime number p. We form the group algebra $\mathbf{Q}[G]$. Let σ be a generator of G. Let

$$e_1 = \frac{1 + \sigma + \sigma^2 + \cdots + \sigma^{p-1}}{p}, \qquad e_2 = 1 - e_1.$$

Then $\tau e_1 = e_1$ for any $\tau \in G$ and consequently $e_1^2 = e_1$. It then follows that $e_2^2 = e_2$ and $e_1 e_2 = 0$. The field $\mathbf{Q}e_1$ is isomorphic to $\mathbf{Q}$. Let $\omega = \sigma e_2$. Then $\omega^p = e_2$. Let $\mathbf{Q}_2 = \mathbf{Q}e_2$. Since $\omega \neq e_2$, and satisfies the irreducible equation

$$X^{p-1} + \cdots + 1 = 0$$

over $\mathbf{Q}_2$, it follows that $\mathbf{Q}_2(\omega)$ is isomorphic to the field obtained by adjoining a primitive p-th root of unity to the rationals. Consequently, $\mathbf{Q}[G]$ admits the direct product decomposition

$$\mathbf{Q}[G] \approx \mathbf{Q} \times \mathbf{Q}(\zeta)$$

where ζ is a primitive p-th root of unity.

As another example, let G be any finite group, and let

$$e_1 = \frac{1}{n} \sum_{\sigma \in G} \sigma.$$

Then for any $\tau \in G$ we have $\tau e_1 = e_1$, and $e_1^2 = e_1$. If we let $e_1' = 1 - e_1$ then $e_1'^2 = e_1'$, and $e_1' e_1 = e_1 e_1' = 0$. Thus for any field k (whose characteristic does not divide the order of G according to conventions in force), we see that

$$k[G] = ke_1 \times k[G]e_1'$$

is a direct product decomposition. In particular, the representation of G on the group algebra $k[G]$ itself contains a 1-dimensional representation on the component ke_1, whose character is the trivial character.

§3. 1-DIMENSIONAL REPRESENTATIONS

By abuse of language, even in characteristic $p > 0$, we say that a **character is 1-dimensional** if it is a homomorphism $G \to k^*$.

Assume that E is a 1-dimensional vector space over k. Let

$$\rho : G \to \mathrm{Aut}_k(E)$$

be a representation. Let $\{v\}$ be a basis of E over k. Then for each $\sigma \in G$, we have

$$\sigma v = \chi(\sigma) v$$

for some element $\chi(\sigma) \in k$, and $\chi(\sigma) \neq 0$ since σ induces an automorphism of E. Then for $\tau \in G$,

$$\tau\sigma v = \chi(\sigma)\tau v = \chi(\sigma)\chi(\tau)v = \chi(\sigma\tau)v.$$

We see that $\chi : G \to k^*$ is a homomorphism, and that our 1-dimensional character is the same type of thing that occurred in Artin's theorem in Galois theory.

Conversely, let $\chi : G \to k^*$ be a homomorphism. Let E be a 1-dimensional k-space, with basis $\{v\}$, and define $\sigma(av) = a\chi(\sigma)v$ for all $a \in k$. Then we see at once that this operation of G on E gives a representation of G, whose associated character is χ.

Since G is finite, we note that

$$\chi(\sigma)^n = \chi(\sigma^n) = \chi(1) = 1.$$

Hence the values of 1-dimensional characters are n-th roots of unity. The 1-dimensional characters form a group under multiplication, and when G is a finite abelian group, we have determined its group of 1-dimensional characters in Chapter I, §9.

Theorem 3.1. *Let G be a finite abelian group, and assume that k is algebraically closed. Then every simple representation of G is 1-dimensional. The simple characters of G are the homomorphisms of G into k^*.*

Proof. The group ring $k[G]$ is semisimple, commutative, and is a direct product of simple rings. Each simple ring is a ring of matrices over k (by Corollary 3.6 Chapter XVII), and can be commutative if and only if it is equal to k.

For every 1-dimensional character χ of G we have

$$\chi(\sigma)^{-1} = \chi(\sigma^{-1}).$$

If k is the field of complex numbers, then

$$\overline{\chi(\sigma)} = \chi(\sigma)^{-1} = \chi(\sigma^{-1}).$$

Corollary 3.2. *Let k be algebraically closed. Let G be a finite group. For any character χ and $\sigma \in G$, the value $\chi(\sigma)$ is equal to a sum of roots of unity with integer coefficients (i.e. coefficients in $\mathbf{Z}$ or $\mathbf{Z}/p\mathbf{Z}$ depending on the characteristic of k).*

Proof. Let H be the subgroup generated by σ. Then H is a cyclic subgroup. A representation of G having character χ can be viewed as a representation for H by restriction, having the same character. Thus our assertion follows from Theorem 3.1.

§4. THE SPACE OF CLASS FUNCTIONS

By a **class function** of G (over k, or with values in k), we shall mean a function $f: G \to k$ such that $f(\sigma\tau\sigma^{-1}) = f(\tau)$ for all $\sigma, \tau \in G$. It is clear that characters are class functions, because for square matrices M, M' we have

$$\operatorname{tr}(MM'M^{-1}) = \operatorname{tr}(M').$$

Thus a class function may be viewed as a function on conjugacy classes.

We shall always extend the domain of definition of a class function to the group ring, by linearity. If

$$\alpha = \sum_{\sigma \in G} a_\sigma \sigma,$$

and f is a class function, we define

$$f(\alpha) = \sum_{\sigma \in G} a_\sigma f(\sigma).$$

Let $\sigma_0 \in G$. If $\sigma \in G$, we write $\sigma \sim \sigma_0$ if σ is conjugate to σ_0, that is, if there exists an element τ such that $\sigma_0 = \tau\sigma\tau^{-1}$. An element of the group ring of type

$$\gamma = \sum_{\sigma \sim \sigma_0} \sigma$$

will also be called a **conjugacy class**.

Proposition 4.1. *An element of $k[G]$ commutes with every element of G if and only if it is a linear combination of conjugacy classes with coefficients in k.*

Proof. Let $\alpha = \sum_{\sigma \in G} a_\sigma \sigma$ and assume $\alpha\tau = \tau\alpha$ for all $\tau \in G$. Then

$$\sum_{\sigma \in G} a_\sigma \tau\sigma\tau^{-1} = \sum_{\sigma \in G} a_\sigma \sigma.$$

Hence $a_{\sigma_0} = a_\sigma$ whenever σ is conjugate to σ_0, and this means that we can write

$$\alpha = \sum_\gamma a_\gamma \gamma$$

where the sum is taken over all conjugacy classes γ.

Remark. We note that the conjugacy classes in fact form a basis of the center of $\mathbf{Z}[G]$ over $\mathbf{Z}$, and thus play a universal role in the theory of representations.

We observe that the conjugacy classes are linearly independent over k, and form a basis for the center of $k[G]$ over k.

Assume for the rest of this section that k is algebraically closed. Then

$$k[G] = \prod_{i=1}^{s} R_i$$

is a direct product of simple rings, and each R_i is a matrix algebra over k. In a direct product, the center is obviously the product of the centers of each factor. Let us denote by k_i the image of k in R_i, in other words,

$$k_i = ke_i,$$

where e_i is the unit element of R_i. Then the center of $k[G]$ is also equal to

$$\prod_{i=1}^{s} k_i$$

which is s-dimensional over k.

If L_i is a typical simple left ideal of R_i, then

$$R_i \approx \mathrm{End}_k(L_i).$$

We let

$$d_i = \dim_k L_i.$$

Then

$$\boxed{d_i^2 = \dim_k R_i \quad \text{and} \quad \sum_{i=1}^{s} d_i^2 = n.}$$

We also have the direct sum decomposition

$$R_i \approx L_i^{(d_i)}$$

as a (G, k)-space.

The above notation will remain fixed from now on.

We can summarize some of our results as follows.

Proposition 4.2. *Let k be algebraically closed. Then the number of conjugacy classes of G is equal to the number of simple characters of G, both of these being equal to the number s above. The conjugacy classes $\gamma_1, \ldots, \gamma_s$ and the unit elements $e_1, \ldots, e_s$ form bases of the center of $k[G]$.*

The number of elements in γ_i will be denoted by h_i. The number of elements in a conjugacy class γ will be denoted by h_γ. We call it the **class number**. The center of the group algebra will be denoted by $Z_k(G)$.

We can view $k[G]$ as a G-module. Its character will be called the **regular character**, and will be denoted by χ_{reg} or r_G if we need to specify the dependence on G. The representation on $k[G]$ is called the **regular representation**. From our direct sum decomposition of $k[G]$ we get

$$\boxed{\chi_{\text{reg}} = \sum_{i=1}^{s} d_i \chi_i.}$$

We shall determine the values of the regular character.

Proposition 4.3. *Let χ_{reg} be the regular character. Then*

$$\chi_{\text{reg}}(\sigma) = 0 \quad \textit{if} \quad \sigma \in G, \sigma \neq 1$$

$$\chi_{\text{reg}}(1) = n.$$

Proof. Let $1 = \sigma_1, \ldots, \sigma_n$ be the elements of G. They form a basis of $k[G]$ over k. The matrix of 1 is the unit $n \times n$ matrix. Thus our second assertion follows. If $\sigma \neq 1$, then multiplication by σ permutes $\sigma_1, \ldots, \sigma_n$, and it is immediately clear that all diagonal elements in the matrix representing σ are 0. This proves what we wanted.

We observe that we have two natural bases for the center $Z_k(G)$ of the group ring. First, the conjugacy classes of elements of G. Second, the elements $e_1, \ldots, e_s$ (i.e. the unit elements of the rings R_i). We wish to find the relation between these, in other words, we wish to find the coefficients of e_i when expressed in terms of the group elements. The next proposition does this. The values of these coefficients will be interpreted in the next section as scalar products. This will clarify their mysterious appearance.

Proposition 4.4. *Assume again that k is algebraically closed. Let*

$$e_i = \sum_{\tau \in G} a_\tau \tau, \qquad a_\tau \in k.$$

Then

$$a_\tau = \frac{1}{n} \chi_{\text{reg}}(e_i \tau^{-1}) = \frac{d_i}{n} \chi_i(\tau^{-1}).$$

Proof. We have for all $\tau \in G$:

$$\chi_{\text{reg}}(e_i \tau^{-1}) = \chi_{\text{reg}}\left(\sum_{\sigma \in G} a_\sigma \sigma \tau^{-1}\right) = \sum_{\sigma \in G} a_\sigma \chi_{\text{reg}}(\sigma \tau^{-1}).$$

By Proposition 4.3, we find

$$\chi_{\text{reg}}(e_i\tau^{-1}) = na_\tau.$$

On the other hand,

$$\chi_{\text{reg}}(e_i\tau^{-1}) = \sum_{j=1}^{s} d_j\chi_j(e_i\tau^{-1}) = d_i\chi_i(e_i\tau^{-1}) = d_i\chi_i(\tau^{-1}).$$

Hence

$$d_i\chi_i(\tau^{-1}) = na_\tau$$

for all $\tau \in G$. This proves our proposition.

Corollary 4.5. *Each e_i can be expressed in terms of group elements with coefficients which lie in the field generated over the prime field by m-th roots of unity, if m is an exponent for G.*

Corollary 4.6. *The dimensions d_i are not divisible by the characteristic of k.*

Proof. Otherwise, $e_i = 0$, which is impossible.

Corollary 4.7. *The simple characters $\chi_1, \ldots, \chi_s$ are linearly independent over k.*

Proof. The proof in Corollary 2.4 applies, since we now know that the characteristic does not divide d_i.

Corollary 4.8. *Assume in addition that k has characteristic 0. Then $d_i | n$ for each i.*

Proof. Multiplying our expression for e_i by n/d_i, and also by e_i, we find

$$\frac{n}{d_i} e_i = \sum_{\sigma \in G} \chi_i(\sigma^{-1})\sigma e_i.$$

Let ζ be a primitive m-th root of unity, and let M be the module over $\mathbf{Z}$ generated by the finite number of elements $\zeta^\nu \sigma e_i$ ($\nu = 0, \ldots, m-1$ and $\sigma \in G$). Then from the preceding relation, we see at once that multiplication by n/d_i maps M into itself. By definition, we conclude that n/d_i is integral over $\mathbf{Z}$, and hence lies in $\mathbf{Z}$, as desired.

Theorem 4.9. *Let k be algebraically closed. Let $Z_k(G)$ be the center of $k[G]$, and let $X_k(G)$ be the k-space of class functions on G. Then $Z_k(G)$ and $X_k(G)$ are the dual spaces of each other, under the pairing*

$$(f, \alpha) \mapsto f(\alpha).$$

The simple characters and the unit elements $e_1, \ldots, e_s$ form orthogonal bases to each other. We have

$$\chi_i(e_j) = \delta_{ij} d_i.$$

Proof. The formula has been proved in the proof of Theorem 2.3. The two spaces involved here both have dimension s, and $d_i \neq 0$ in k. Our proposition is then clear.

§5. ORTHOGONALITY RELATIONS

Throughout this section, we assume that k is algebraically closed.

If R is a subring of k, we denote by $X_R(G)$ the R-module generated over R by the characters of G. It is therefore the module of functions which are linear combinations of simple characters with coefficients in R. If R is the prime ring (i.e. the integers $\mathbf{Z}$ or the integers mod p if k has characteristic p), then we denote $X_R(G)$ by $X(G)$.

We shall now define a bilinear map on $X(G) \times X(G)$. If $f, g \in X(G)$, we define

$$\langle f, g \rangle = \frac{1}{n} \sum_{\sigma \in G} f(\sigma) g(\sigma^{-1}).$$

Theorem 5.1. *The symbol $\langle f, g \rangle$ for $f, g \in X(G)$ takes on values in the prime ring. The simple characters form an orthonormal basis for $X(G)$, in other words*

$$\langle \chi_i, \chi_j \rangle = \delta_{ij}.$$

For each ring $R \subset k$, the symbol has a unique extension to an R-bilinear form $X_R(G) \times X_R(G) \to R$, given by the same formula as above.

Proof. By Proposition 4.4, we find

$$\chi_j(e_i) = \frac{d_i}{n} \sum_{\sigma \in G} \chi_i(\sigma^{-1}) \chi_j(\sigma).$$

If $i \neq j$ we get 0 on the left-hand side, so that χ_i and χ_j are orthogonal. If $i = j$ we get d_i on the left-hand side, and we know that $d_i \neq 0$ in k, by Corollary 4.6. Hence $\langle \chi_i, \chi_i \rangle = 1$. Since every element of $X(G)$ is a linear combination of simple characters with integer coefficients, it follows that the values of our bilinear map are in the prime ring. The extension statement is obvious, thereby proving our theorem.

Assume that k has characteristic 0. Let m be an exponent for G, and let R contain the m-th roots of unity. If R has an automorphism of order 2 such that its effect on a root of unity is $\zeta \mapsto \zeta^{-1}$, then we shall call such an automorphism a **conjugation**, and denote it by $a \mapsto \bar{a}$.

Theorem 5.2. *Let k have characteristic 0, and let R be a subring containing the m-th roots of unity, and having a conjugation. Then the bilinear form on $X(G)$ has a unique extension to a hermitian form*

$$X_R(G) \times X_R(G) \to R,$$

given by the formula

$$\langle f, g \rangle = \frac{1}{n} \sum_{\sigma \in G} f(\sigma)\overline{g(\sigma)}.$$

The simple characters constitute an orthonormal basis of $X_R(G)$ with respect to this form.

Proof. The formula given in the statement of the theorem gives the same value as before for the symbol $\langle f, g \rangle$ when f, g lie in $X(G)$. Thus the extension exists, and is obviously unique.

We return to the case when k has arbitrary characteristic.

Let $Z(G)$ denote the additive group generated by the conjugacy classes $\gamma_1, \dots, \gamma_s$ over the prime ring. It is of dimension s. We shall define a bilinear map on $Z(G) \times Z(G)$. If $\alpha = \sum a_\sigma \sigma$ has coefficients in the prime ring, we denote by α^- the element $\sum a_\sigma \sigma^{-1}$.

Proposition 5.3. *For $\alpha, \beta \in Z(G)$, we can define a symbol $\langle \alpha, \beta \rangle$ by either one of the following expressions, which are equal:*

$$\langle \alpha, \beta \rangle = \frac{1}{n} \chi_{\text{reg}}(\alpha\beta^-) = \frac{1}{n} \sum_{\nu=1}^{s} \chi_\nu(\alpha)\chi_\nu(\beta^-).$$

The values of the symbol lie in the prime ring.

Proof. Each expression is linear in its first and second variable. Hence to prove their equality, it will suffice to prove that the two expressions are equal when we replace α by e_i and β by an element τ of G. But then, our equality is equivalent to

$$\chi_{\text{reg}}(e_i \tau^{-1}) = \sum_{\nu=1}^{s} \chi_\nu(e_i)\chi_\nu(\tau^{-1}).$$

Since $\chi_v(e_i) = 0$ unless $v = i$, we see that the right-hand side of this last relation is equal to $d_i\chi_i(\tau^{-1})$. Our two expressions are equal in view of Proposition 4.4.

The fact that the values lie in the prime ring follows from Proposition 4.3: The values of the regular character on group elements are equal to 0 or n, and hence in characteristic 0, are integers divisible by n.

As with $X_R(G)$, we use the notation $Z_R(G)$ to denote the R-module generated by $\gamma_1, \ldots, \gamma_s$ over an arbitrary subring R of k.

Lemma 5.4. *For each ring R contained in k, the pairing of Proposition 5.3 has a unique extension to a map*

$$Z_R(G) \times Z(G) \to R$$

which is R-linear in its first variable. If R contains the m-th roots of unity, where m is an exponent for G, and also contains $1/n$, then $e_i \in Z_R(G)$ for all i. The class number h_i is not divisible by the characteristic of k, and we have

$$e_i = \sum_{\nu=1}^{s} \langle e_i, \gamma_\nu \rangle \frac{1}{h_\nu} \gamma_\nu .$$

Proof. We note that h_i is not divisible by the characteristic because it is the index of a subgroup of G (the isotropy group of an element in γ_i when G operates by conjugation), and hence h_i divides n. The extension of our pairing as stated is obvious, since $\gamma_1, \ldots, \gamma_s$ form a basis of $Z(G)$ over the prime ring. The expression of e_i in terms of this basis is only a reinterpretation of Proposition 4.4 in terms of the present pairing.

Let E be a free module over a subring R of k, and assume that we have a bilinear symmetric (or hermitian) form on E. Let $\{v_1, \ldots, v_s\}$ be an orthogonal basis for this module. If

$$v = a_1 v_1 + \cdots + a_s v_s$$

with $a_i \in R$, then we call $a_1, \ldots, a_s$ the **Fourier coefficients** of v with respect to our basis. In terms of the form, these coefficients are given by

$$a_i = \frac{\langle v, v_i \rangle}{\langle v_i, v_i \rangle}$$

provided $\langle v_i, v_i \rangle \neq 0$.

We shall see in the next theorem that the expression for e_i in terms of $\gamma_1, \ldots, \gamma_s$ is a Fourier expansion.

Theorem 5.5. *The conjugacy classes $\gamma_1, \ldots, \gamma_s$ constitute an orthogonal basis for $Z(G)$. We have $\langle \gamma_i, \gamma_i \rangle = h_i$. For each ring R contained in k, the bilinear map of Proposition 5.3 has a unique extension to a R-bilinear map*

$$Z_R(G) \times Z_R(G) \to R.$$

Proof. We use the lemma. By linearity, the formula in the lemma remains valid when we replace R by k, and when we replace e_i by any element of $Z_k(G)$, in particular when we replace e_i by γ_i. But $\{\gamma_1, \dots, \gamma_s\}$ is a basis of $Z_k(G)$, over k. Hence we find that $\langle \gamma_i, \gamma_i \rangle = h_i$ and $\langle \gamma_i, \gamma_j \rangle = 0$ if $i \neq j$, as was to shown.

Corollary 5.6. *If G is commutative, then*

$$\frac{1}{n} \sum_{\nu=1}^{n} \chi_\nu(\sigma)\chi_\nu(\tau^{-1}) = \begin{cases} 0 & \text{if } \sigma \text{ is not equal to } \tau \\ 1 & \text{if } \sigma \text{ is equal to } \tau. \end{cases}$$

Proof. When G is commutative, each conjugacy class has exactly one element, and the number of simple characters is equal to the order of the group.

We consider the case of characteristic 0 for our $Z(G)$ just as we did for $X(G)$. Let k have characteristic 0, and R be a subring of k containing the m-th roots of unity, and having a conjugation. Let $\alpha = \sum_{\sigma \in G} a_\sigma \sigma$ with $a_\sigma \in R$. We define

$$\bar{\alpha} = \sum_{\sigma \in G} \bar{a}_\sigma \sigma^{-1}.$$

Theorem 5.7. *Let k have characteristic 0, and let R be a subring of k, containing the m-th roots of unity, and having a conjugation. Then the pairing of Proposition 5.3 has a unique extension to a hermitian form*

$$Z_R(G) \times Z_R(G) \to R$$

given by the formulas

$$\langle \alpha, \beta \rangle = \frac{1}{n} \chi_{\text{reg}}(\alpha\bar{\beta}) = \frac{1}{n} \sum_{\nu=1}^{s} \chi_\nu(\alpha)\overline{\chi_\nu(\beta)}.$$

The conjugacy classes $\gamma_1, \dots, \gamma_s$ form an orthogonal basis for $Z_R(G)$. If R contains $1/n$, then $e_1, \dots, e_s$ lie in $Z_R(G)$ and also form an orthogonal basis for $Z_R(G)$. We have $\langle e_i, e_i \rangle = d_i^2/n$.

Proof. The formula given in the statement of the theorem gives the same value as the symbol $\langle \alpha, \beta \rangle$ of Proposition 5.3 when α, β lie in $Z(G)$. Thus the extension exists, and is obviously unique. Using the second formula in Proposition 5.3, defining the scalar product, and recalling that $\chi_\nu(e_i) = 0$ if $\nu \neq i$, we see that

$$\langle e_i, e_i \rangle = \frac{1}{n} \chi_i(e_i)\overline{\chi_i(e_i)},$$

whence our assertion follows.

We observe that the Fourier coefficients of e_i relative to the basis $\gamma_1, \dots, \gamma_s$ are the same with respect to the bilinear form of Theorem 5.5, or the hermitian form of Theorem 5.7. This comes from the fact that $\gamma_1, \dots, \gamma_s$ lie in $Z(G)$, and form a basis of $Z(G)$ over the prime ring.

We shall now reprove and generalize the orthogonality relations by another method. Let E be a finite dimensional (G, k)-space, so we have a representation

$$G \to \mathrm{Aut}_k(E).$$

After selecting a basis of E, we get a representation of G by $d \times d$ matrices. If $\{v_1, \dots, v_d\}$ is the basis, then we have the dual basis $\{\lambda_1, \dots, \lambda_d\}$ such that $\lambda_i(v_j) = \delta_{ij}$. If an element σ of G is represented by a matrix $(\rho_{ij}(\sigma))$, then each coefficient $\rho_{ij}(\sigma)$ is a function of σ, called the ***ij*-coefficient function**. We can also write

$$\rho_{ij}(\sigma) = \lambda_j(\sigma v_i).$$

But instead of indexing elements of a basis or the dual basis, we may just as well work with any functional λ on E, and any vector v. Then we get a function

$$\sigma \mapsto \lambda(\sigma v) = \rho_{\lambda, v}(\sigma),$$

which will also be called a **coefficient function**. In fact, one can always complete $v = v_1$ to a basis such that $\lambda = \lambda_1$ is the first element in the dual basis, but using the notation $\rho_{\lambda, v}$ is in many respects more elegant.

We shall constantly use:

Schur's Lemma. *Let E, F be simple (G, k)-spaces, and let*

$$\varphi : E \to F$$

be a homomorphism. Then either $\varphi = 0$ or φ is an isomorphism.

Proof. Indeed, the kernel of φ and the image of φ are subspaces, so the assertion is obvious.

We use the same formula as before to define a scalar product on the space of all k-valued functions on G, namely

$$\langle f, g \rangle = \frac{1}{n} \sum_{\sigma \in G} f(\sigma) g(\sigma^{-1}).$$

We shall derive various orthogonality relations among coefficient functions.

Theorem 5.8. *Let E, F be simple (G, k)-spaces. Let λ be a k-linear functional on E, let $x \in E$ and $y \in F$. If E, F are not isomorphic, then*

$$\sum_{\sigma \in G} \lambda(\sigma x) \sigma^{-1} y = 0.$$

If μ is a functional on F then the coefficient functions $\rho_{\lambda,x}$ and $\rho_{\mu,y}$ are orthogonal, that is

$$\sum_{\sigma \in G} \lambda(\sigma x)\mu(\sigma^{-1} y) = 0.$$

Proof. The map $x \mapsto \sum \lambda(\sigma x)\sigma^{-1} y$ is a G-homomorphism of E into F, so Schur's lemma concludes the proof of the first statement. The second comes by applying the functional μ.

As a corollary, we see that if χ, ψ are distinct irreducible characters of G over k, then

$$\langle \chi, \psi \rangle = 0,$$

that is the characters are orthogonal. Indeed, the character associated with a representation ρ is the sum of the diagonal coefficient functions,

$$\chi = \sum_{i=1}^{d} \rho_{ii},$$

where d is the dimension of the representation. Two distinct characters correspond to non-isomorphic representations, so we can apply Proposition 5.8.

Lemma 5.9. *Let E be a simple (G, k)-space. Then any G-endomorphism of E is equal to a scalar multiple of the identity.*

Proof. The algebra $\text{End}_{G,k}(E)$ is a division algebra by Schur's lemma, and is finite dimensional over k. Since k is assumed algebraically closed, it must be equal to k because any element generates a commutative subfield over k. This proves the lemma.

Lemma 5.10. *Let E be a representation space for G of dimension d. Let λ be a functional on E, and let $x \in E$. Let $\varphi_{\lambda,x} \in \text{End}_k(E)$ be the endomorphism such that*

$$\varphi_{\lambda,x}(y) = \lambda(y)x.$$

Then $\text{tr}(\varphi_{\lambda,x}) = \lambda(x)$.

Proof. If $x = 0$ the statement is obvious. Let $x \neq 0$. If $\lambda(x) \neq 0$ we pick a basis of E consisting of x and a basis of the kernel of λ. If $\lambda(x) = 0$, we pick a basis of E consisting of a basis for the kernel of λ, and one other element. In either case it is immediate from the corresponding matrix representing $\varphi_{\lambda,x}$ that the trace is given by the formula as stated in the lemma.

Theorem 5.11. *Let $\rho: G \to \text{Aut}_k(E)$ be a simple representation of G, of dimension d. Then the characteristic of k does not divide d. Let $x, y \in E$. Then for any functionals λ, μ on E,*

$$\sum_{\sigma \in G} \lambda(\sigma x)\mu(\sigma^{-1} y) = \frac{n}{d} \lambda(y)\mu(x).$$

Proof. It suffices to prove that

$$\sum_{\sigma \in G} \lambda(\sigma x)\sigma^{-1}y = \frac{n}{d}\lambda(y)x.$$

For fixed y the map

$$x \mapsto \sum_{\sigma \in G} \lambda(\sigma x)\sigma^{-1}y$$

is immediately verified to be a G-endomorphism of E, so is equal to cI for some $c \in k$ by Lemma 5.9. In fact, it is equal to

$$\sum_{\sigma \in G} \rho(\sigma^{-1}) \circ \varphi_{\lambda, y} \circ \rho(\sigma).$$

The trace of this expression is equal to $n \cdot \operatorname{tr}(\varphi_{\lambda, y})$ by Lemma 5.10, and also to dc. Taking λ, y such that $\lambda(y) = 1$ shows that the characteristic does not divide d, and then we can solve for c as stated in the theorem.

Corollary 5.12. *Let χ be the character of the representation of G on the simple space E. Then*

$$\langle \chi, \chi \rangle = 1.$$

Proof. This follows immediately from the theorem, and the expression of χ as

$$\chi = \rho_{11} + \cdots + \rho_{dd}.$$

We have now recovered the fact that the characters of simple representations are orthonormal. We may then recover the idempotents in the group ring, that is, if $\chi_1, \ldots, \chi_s$ are the simple characters, we may now *define*

$$e_i = \frac{d_i}{n} \sum_{\sigma \in G} \chi_i(\sigma)\sigma^{-1}.$$

Then the orthonormality of the characters yields the formulas:

Corollary 5.13. $\chi_i(e_j) = \delta_{ij} d_i$ *and* $\chi_{\text{reg}} = \sum_{i=1}^{s} d_i \chi_i$.

Proof. The first formula is a direct application of the orthonormality of the characters. The second formula concerning the regular character is obtained by writing

$$\chi_{\text{reg}} = \sum_j m_j \chi_j$$

with unknown coefficients. We know the values $\chi_{\text{reg}}(1) = n$ and $\chi_{\text{reg}}(\sigma) = 0$ if $\sigma \neq 1$. Taking the scalar product of χ_{reg} with χ_i for $i = 1, \dots, s$ immediately yields the desired values for the coefficients m_j.

Since a character is a class function, one sees directly that each e_i is a linear combination of conjugacy classes, and so is in the center of the group ring $k[G]$.

Now let E_i be a representation space of χ_i, and let ρ_i be the representation of G or $k[G]$ on E_i. For $\alpha \in k[G]$ we let $\rho_i(\alpha): E_i \to E_i$ be the map such that $\rho_i(\alpha)x = \alpha x$ for all $x \in E_i$.

Proposition 5.14. *We have*

$$\rho_i(e_i) = \text{id} \quad \textit{and} \quad \rho_i(e_j) = 0 \quad \textit{if } i \neq j.$$

Proof. The map $x \mapsto e_i x$ is a G-homomorphism of E_i into itself since e_i is in the center of $k[G]$. Hence by Lemma 5.9 this homomorphism is a scalar multiple of the identity. Taking the trace and using the orthogonality relations between simple characters immediately gives the desired value of this scalar.

We now find that

$$\sum_{i=1}^{s} e_i = 1$$

because the group ring $k[G]$ is a direct sum of simple spaces, possibly with multiplicities, and operates faithfully on itself.

The orthonormality relations also allow us to expand a function in a Fourier expression, relative to the characters if it is a class function, and relative to the coefficient functions in general. We state this in two theorems.

Theorem 5.15. *Let f be a class function on G. Then*

$$f = \sum_{i=1}^{s} \langle f, \chi_i \rangle \chi_i.$$

Proof. The number of conjugacy class is equal to the number of distinct characters, and these are linearly independent, so they form a basis for the class functions. The coefficients are given by the stated formula, as one sees by taking the scalar product of f with any character χ_j and using the orthonormality.

Theorem 5.16. *Let $\rho^{(i)}$ be a matrix representation of G on E_i relative to a choice of basis, and let $\rho^{(i)}_{\nu,\mu}$ be the coefficient functions of this matrix, $i = 1, \dots, s$ and $\nu, \mu = 1, \dots, d_i$. Then the functions $\rho^{(i)}_{\nu,\mu}$ form an orthogonal basis for the space of all functions on G, and hence for any function f on G we have*

$$f = \sum_{i=1}^{s} \sum_{\nu,\mu} \frac{1}{d_i} \langle f, \rho^{(i)}_{\nu,\mu} \rangle \rho^{(i)}_{\nu,\mu}.$$

Proof. That the coefficient functions form an orthogonal basis follows from Theorems 5.8 and 5.11. The expression of f in terms of this basis is then merely the standard Fourier expansion relative to any scalar product. This concludes the proof.

Suppose now for concreteness that $k = \mathbf{C}$ is the complex numbers. Recall that an **effective character** χ is an element of $X(G)$, such that if

$$\chi = \sum_{i=1}^{s} m_i \chi_i$$

is a linear combination of the simple characters with integral coefficients, then we have $m_i \geqq 0$ for all i. In light of the orthonormality of the simple characters, we get for all elements $\chi \in X(G)$ the relations

$$\|\chi\|^2 = \langle \chi, \chi \rangle = \sum_{i=1}^{s} m_i^2 \quad \text{and} \quad m_i = \langle \chi, \chi_i \rangle.$$

Hence we get (a) of the next theorem.

Theorem 5.17. **(a)** *Let χ be an effective character in $X(G)$. Then χ is simple over* $\mathbf{C}$ *if and only if $\|\chi\|^2 = 1$, or alternatively,*

$$\sum_{\sigma \in G} |\chi(\sigma)|^2 = \#(G).$$

(b) *Let χ, ψ be effective characters in $X(G)$, and let E, F be their representation spaces over* $\mathbf{C}$. *Then*

$$\langle \chi, \psi \rangle_G = \dim \operatorname{Hom}_G(E, F).$$

Proof. The first part has been proved, and for (b), let $\psi = \sum q_i \chi_i$. Then by orthonormality, we get

$$\langle \chi, \psi \rangle_G = \sum m_i q_i.$$

But if E_i is the representation space of χ_i over $\mathbf{C}$, then by Schur's lemma

$$\dim \operatorname{Hom}_G(E_i, E_i) = 1 \text{ and } \dim \operatorname{Hom}_G(E_i, E_j) = 0 \text{ for } i \neq j.$$

Hence $\dim \operatorname{Hom}_G(E, F) = \sum m_i q_i$, thus proving (b).

Corollary 5.18 *With the above notation and $k = \mathbf{C}$ for simplicity, we have:*
(a) *The multiplicity of 1_G in $E^\vee \otimes F$ is* $\dim_k \operatorname{inv}_G(E^\vee \otimes F)$.
(b) *The (G, k)-space E is simple if and only if 1_G has multiplicity* 1 *in $E^\vee \otimes E$.*

Proof. Immediate from Theorem 5.17 and formula (3) of §1.

Remark. The criterion of Theorem 5.17(a) is useful in testing whether a representation is simple. In practice, representations are obtained by inducing from 1-dimensional characters, and such induced representations do have a tendency to be irreducible. We shall see a concrete case in §12.

§6. INDUCED CHARACTERS

The notation is the same as in the preceding section. However, we don't need all the results proved there; all we need is the bilinear pairing on $X(G)$, and its extension to

$$X_R(G) \times X_R(G) \to R.$$

The symbol $\langle\ ,\ \rangle$ may be interpreted either as the bilinear extension, or the hermitian extension according to Theorem 5.2.

Let S be a subgroup of G. We have an R-linear map called the restriction

$$\operatorname{res}_S^G : X_R(G) \to X_R(S)$$

which to each class function on G associates its restriction to S. It is a ring-homomorphism. We sometimes let f_S denote the restriction of f to S.

We shall define a map in the opposite direction,

$$\operatorname{ind}_S^G : X_R(S) \to X_R(G),$$

which we call the **induction map**. If $g \in X_R(S)$, we extend g to g_S on G by letting $g_S(\sigma) = 0$ if $\sigma \notin S$. Then we define the **induced function**

$$g^G(\sigma) = \operatorname{ind}_S^G(g)(\sigma) = \frac{1}{(S:1)} \sum_{\tau \in G} g_S(\tau\sigma\tau^{-1}).$$

Then $\operatorname{ind}_S^G(g)$ is a class function on G. It is clear that ind_S^G is R-linear.

Since we deal with two groups S and G, we shall denote the scalar product by $\langle\ ,\ \rangle_S$ and $\langle\ ,\ \rangle_G$ when it is taken with these respective groups. The next theorem shows among other things that the restriction and transfer are adjoint to each other with respect to our form.

Theorem 6.1. *Let S be a subgroup of G. Then the following rules hold:*

(i) (**Frobenius reciprocity**) *For $f \in X_R(G)$, and $g \in X_R(S)$ we have*

$$\langle \operatorname{ind}_S^G(g), f\rangle_G = \langle g, \operatorname{Res}_S^G(f)\rangle_S.$$

(ii) $\operatorname{Ind}_S^G(g)f = \operatorname{ind}_S^G(gf_S)$.

(iii) *If $T \subset S \subset G$ are subgroups of G, then*

$$\operatorname{ind}_S^G \circ \operatorname{ind}_T^S = \operatorname{ind}_T^G.$$

(iv) *If $\sigma \in G$ and g^σ is defined by $g^\sigma(\tau^\sigma) = g(\tau)$, where $\tau^\sigma = \sigma^{-1}\tau\sigma$, then*

$$\operatorname{ind}_S^G(g) = \operatorname{ind}_{S^\sigma}^G(g^\sigma).$$

(v) *If ψ is an effective character of S then* $\operatorname{ind}_S^G(\psi)$ *is effective.*

Proof. Let us first prove (ii). We must show that $g^G f = (g f_S)^G$. We have

$$(g^G f)(\tau) = \frac{1}{(S:1)} \sum_{\sigma \in G} g_S(\sigma\tau\sigma^{-1}) f(\tau) = \frac{1}{(S:1)} \sum_{\sigma \in G} g_S(\sigma\tau\sigma^{-1}) f(\sigma\tau\sigma^{-1}).$$

The last expression just obtained is equal to $(g f_S)^G$, thereby proving (ii). Let us sum over τ in G. The only non-zero contributions in our double sum will come from those elements of S which can be expressed in the form $\sigma\tau\sigma^{-1}$ with $\sigma, \tau \in G$. The number of pairs (σ, τ) such that $\sigma\tau\sigma^{-1}$ is equal to a fixed element of G is equal to n (because for every $\lambda \in G$, $(\sigma\lambda, \lambda^{-1}\tau\lambda)$ is another such pair, and the total number of pairs is n^2). Hence our expression is equal to

$$(G:1)\frac{1}{(S:1)} \sum_{\lambda \in S} g(\lambda) f(\lambda).$$

Our first rule then follows from the definitions of the scalar products in G and S respectively.

Now let $g = \psi$ be an effective character of S, and let $f = \chi$ be a simple character of G. From (i) we find that the Fourier coefficients of g^G are integers $\geqq 0$ because $\mathrm{res}_S^G(\chi)$ is an effective character of S. Therefore the scalar product

$$\langle \psi, \mathrm{res}_S^G(\chi) \rangle_S$$

is $\geqq 0$. Hence ψ^G is an effective character of G, thereby proving (v).

In order to prove the transitivity property, it is convenient to use the following notation.

Let $\{c\}$ denote the set of *right* cosets of S in G. For each right coset c, we select a fixed coset representative denoted by $\bar{c}$. Thus if $\bar{c}_1, \ldots, \bar{c}_r$ are these representatives, then

$$G = \bigcup_c c = \bigcup_c S\bar{c} = \bigcup_{i=1}^{r} S\bar{c}_i.$$

Lemma 6.2. *Let g be a class function on S. Then*

$$\mathrm{ind}_S^G(g)(\xi) = \sum_{i=1}^{r} g_S(\bar{c}_i \xi \bar{c}_i^{-1}).$$

Proof. We can split the sum over all $\sigma \in G$ in the definition of the induced function into a double sum

$$\sum_{\sigma \in G} = \sum_{\sigma \in S} \sum_{i=1}^{r}$$

and observe that each term $g_S(\sigma\bar{c}\xi\bar{c}^{-1}\sigma^{-1})$ is equal to $g_S(\bar{c}\xi\bar{c}^{-1})$ if $\sigma \in S$, because g is a class function. Hence the sum over $\sigma \in S$ is enough to cancel the factor $1/(S:1)$ in front, to give the expression in the lemma.

If $T \subset S \subset G$ are subgroups of G, and if

$$G = \bigcup S\bar{c}_i \quad \text{and} \quad S = \bigcup T\bar{d}_j$$

are decompositions into right cosets, then $\{\bar{d}_j\bar{c}_i\}$ form a system of representatives for the right cosets of T in G. From this the transitivity property (iii) is obvious. We shall leave (iv) as an exercise (trivial, using the lemma).

§7. INDUCED REPRESENTATIONS

Let G be a group and S a subgroup of finite index. Let F be an S-module. We consider the category $\mathfrak{C}$ whose objects are S-homomorphisms $\varphi : F \to E$ of F into a G-module E. (We note that a G-module E can be regarded as an S-module by restriction.) If $\varphi' : F \to E'$ is another object in $\mathfrak{C}$, we define a morphism $\varphi' \to \varphi$ in $\mathfrak{C}$ to be a G-homomorphism $\eta : E' \to E$ making the following diagram commutative:

$$\begin{array}{ccc} & & E' \\ & \overset{\varphi'}{\nearrow} & \downarrow \eta \\ F & & \\ & \underset{\varphi}{\searrow} & \\ & & E \end{array}$$

A universal object in $\mathfrak{C}$ is determined up to a unique G-isomorphism. It will be denoted by

$$\text{ind}_S^G : F \to \text{ind}_S^G(F).$$

We shall prove below that a universal object always exists. If $\varphi : F \to E$ is a universal object, we call E **an induced module**. It is uniquely determined, up to a unique G-isomorphism making a diagram commutative. For convenience, we shall select one induced module such that φ is an inclusion. We shall then call this particular module $\text{ind}_S^G(F)$ **the** G-module **induced** by F. In particular, given an S-homomorphism φ: $F \to E$ into a G-module E, there is a unique G-homomorphism φ_*: $\text{ind}_S^G(F) \to E$ making the following diagram commutative:

$$\begin{array}{ccc} & & \text{ind}_S^G(F) \\ & \overset{\text{ind}_S^G}{\nearrow} & \downarrow \varphi_* = \text{ind}_S^G(\varphi) \\ F & & \\ & \underset{\varphi}{\searrow} & \\ & & E \end{array}$$

The association $\varphi \mapsto \text{ind}_S^G(\varphi)$ then induces an isomorphism

$$\boxed{\text{Hom}_G(\text{ind}_S^G(F), E) \approx \text{Hom}_S(F, \text{res}_S^G(E)),}$$

for an S-module F and a G-module E. We shall see in a moment that ind_S^G is a functor from Mod(S) to Mod(G), and the above formula may be described as saying that **induction is the adjoint functor of restriction**. One also calls this relation **Frobenius reciprocity for modules**, because Theorem 6.1(i) is a corollary.

Sometimes, if the reference to F as an S-module is clear, we shall omit the subscript S, and write simply

$$\text{ind}^G(F)$$

for the induced module.

Let $f: F' \to F$ be an S-homomorphism. If

$$\varphi_S^G : F' \to \text{ind}_S^G(F')$$

is a G-module induced by F', then there exists a unique G-homomorphism $\text{ind}_S^G(F') \to \text{ind}_S^G(F)$ making the following diagram commutative:

$$\begin{array}{ccc} F' & \xrightarrow{\varphi_S^G} & \text{ind}_S^G(F') \\ {\scriptstyle f}\downarrow & \searrow & \downarrow{\scriptstyle \text{ind}_S^G(f)} \\ F & \xrightarrow[\varphi_S^G]{} & \text{ind}_S^G(F) \end{array}$$

It is simply the G-homomorphism corresponding to the universal property for the S-homomorphism $\varphi_G^S \circ f$, represented by a dashed line in our diagram. *Thus* ind_S^G *is a functor, from the category of S-modules to the category of G-modules.*

From the universality and uniqueness of the induced module, we get some formal properties:

ind_S^G *commutes with direct sums: If we have an S-direct sum* $F \oplus F'$, *then*

$$\text{ind}_S^G(F \oplus F') \approx \text{ind}_S^G(F) \oplus \text{ind}_S^G(F'),$$

the direct sum on the right being a G-direct sum.

If $f, g: F' \to F$ *are S-homomorphisms, then*

$$\text{ind}_S^G(f + g) = \text{ind}_S^G(f) + \text{ind}_S^G(g).$$

If $T \subset S \subset G$ *are subgroups of G, and F is a T-module, then*

$$\text{ind}_S^G \circ \text{ind}_T^S(F) \approx \text{ind}_T^G(F).$$

In all three cases, the equality between the left member and the right member of our equations follows at once by using the uniqueness of the universal object. We shall leave the verifications to the reader.

To prove the existence of the induced module, we let $M_G^S(F)$ be the additive group of functions $f: G \to F$ satisfying

$$\sigma f(\xi) = f(\sigma\xi)$$

for $\sigma \in S$ and $\xi \in G$. We define an operation of G on $M_G^S(F)$ by letting

$$(\sigma f)(\xi) = f(\xi\sigma)$$

for $\sigma, \xi \in G$. It is then clear that $M_G^S(F)$ is a G-module.

Proposition 7.1. *Let $\varphi : F \to M_G^S(F)$ be such that $\varphi(x) = \varphi_x$ is the map*

$$\varphi_x(\tau) = \begin{cases} 0 & \text{if} \quad \tau \notin S \\ \tau x & \text{if} \quad \tau \in S. \end{cases}$$

Then φ is an S-homomorphism, $\varphi : F \to M_G^S(F)$ is universal, and φ is injective. The image of φ consists of those elements $f \in M_G^S(F)$ such that $f(\tau) = 0$ if $\tau \notin S$.

Proof. Let $\sigma \in S$ and $x \in F$. Let $\tau \in G$. Then

$$(\sigma\varphi_x)(\tau) = \varphi_x(\tau\sigma).$$

If $\tau \in S$, then this last expression is equal to $\varphi_{\sigma x}(\tau)$. If $\tau \notin S$, then $\tau\sigma \notin S$, and hence both $\varphi_{\sigma x}(\tau)$ and $\varphi_x(\tau\sigma)$ are equal to 0. Thus φ is an S-homomorphism, and it is immediately clear that φ is injective. Furthermore, if $f \in M_G^S(F)$ is such that $f(\tau) = 0$ if $\tau \notin S$, then from the definitions, we conclude that $f = \varphi_x$ where $x = f(1)$.

There remains to prove that φ is universal. To do this, we shall analyze more closely the structure of $M_G^S(F)$.

Proposition 7.2. *Let $G = \bigcup_{i=1}^{r} S\bar{c}_i$ be a decomposition of G into right cosets. Let F_1 be the additive group of functions in $M_G^S(F)$ having value 0 at elements $\xi \in G$, $\xi \notin S$. Then*

$$M_G^S(F) = \bigoplus_{i=1}^{r} \bar{c}_i^{-1} F_1,$$

the direct sum being taken as an abelian group.

Proof. For each $f \in M_G^S(F)$, let f_i be the function such that

$$f_i(\xi) = \begin{cases} 0 & \text{if} \quad \xi \notin S\bar{c}_i \\ f(\xi) & \text{if} \quad \xi \in S\bar{c}_i. \end{cases}$$

For all $\sigma \in S$ we have $f_i(\sigma \bar{c}_i) = (\bar{c}_i f_i)(\sigma)$. It is immediately clear that $\bar{c}_i f_i$ lies in F_1, and

$$f = \sum_{i=1}^{r} \bar{c}_i^{-1}(\bar{c}_i f_i).$$

Thus $M_G^S(F)$ is the sum of the subgroups $\bar{c}_i^{-1}F_1$. It is clear that this sum is direct, as desired.

We note that $\{\bar{c}_1^{-1}, \ldots, \bar{c}_r^{-1}\}$ form a system of representatives for the *left* cosets of S in G. The operation of G on $M_G^S(F)$ is defined by the presceding direct sum decomposition. We see that G permutes the factors transitively. The factor F_1 is S-isomorphic to the original module F, as stated in Proposition 7.1.

Suppose that instead of considering arbitrary modules, we start with a commutative ring R and consider only R-modules E on which we have a representation of G, i.e. a homomorphism $G \to \mathrm{Aut}_R(E)$, thus giving rise to what we call a (G, R)-module. Then it is clear that all our constructions and definitions can be applied in this context. Therefore if we have a representation of S on an R-module F, then we obtain an induced representation of G on $\mathrm{ind}_S^G(F)$. Then we deal with the category $\mathfrak{C}$ of S-homomorphisms of an (S, R)-module into a (G, R)-module. To simplify the notation, we may write "G-module" to mean "(G, R)-module" when such a ring R enters as a ring of coefficients.

Theorem 7.3. *Let $\{\lambda_1, \ldots, \lambda_r\}$ be a system of left coset representatives of S in G. There exists a G-module E containing F as an S-submodule, such that*

$$E = \bigoplus_{i=1}^{r} \lambda_i F$$

is a direct sum (as R-modules). Let $\varphi : F \to E$ be the inclusion mapping. Then φ is universal in our category $\mathfrak{C}$, i.e. E is an induced module.

Proof. By the usual set-theoretic procedure of replacing F_1 by F in $M_G^S(F)$, obtain a G-module E containing F as a S-submodule, and having the desired direct sum decomposition. Let $\varphi' : F \to E'$ be an S-homomorphism into a G-module E'. We define

$$h : E \to E'$$

by the rule

$$h(\lambda_1 x_1 + \cdots + \lambda_r x_r) = \lambda_1 \varphi'(x_1) + \cdots + \lambda_r \varphi'(x_r)$$

for $x_i \in F$. This is well defined since our sum for E is direct. We must show that h is a G-homomorphism. Let $\sigma \in G$. Then

$$\sigma \lambda_i = \lambda_{\sigma(i)} \tau_{\sigma, i}$$

where $\sigma(i)$ is some index depending on σ and i, and $\tau_{\sigma, i}$ is an element of S, also

depending on σ, i. Then

$$h(\sigma\lambda_i x_i) = h(\lambda_{\sigma(i)}\tau_{\sigma,i}x_i) = \lambda_{\sigma(i)}\varphi'(\tau_{\sigma,i}x_i).$$

Since φ' is an S-homomorphism, we see that this expression is equal to

$$\lambda_{\sigma(i)}\tau_{\sigma,i}\varphi'(x_i) = \sigma h(\lambda_i x_i).$$

By linearity, we conclude that h is a G-homomorphism, as desired.

In the next proposition we return to the case when R is our field k.

Proposition 7.4. *Let ψ be the character of the representation of S on the k-space F. Let E be the space of an induced representation. Then the character χ of E is equal to the induced character ψ^G, i.e. is given by the formula*

$$\chi(\xi) = \sum_c \psi_0(\bar{c}\xi\bar{c}^{-1}),$$

where the sum is taken over the right cosets c of S in G, $\bar{c}$ is a fixed coset representative for c, and ψ_0 is the extension of ψ to G obtained by setting $\psi_0(\sigma) = 0$ if $\sigma \notin S$.

Proof. Let $\{w_1, \ldots, w_m\}$ be a basis for F over k. We know that

$$E = \bigoplus \bar{c}^{-1}F.$$

Let σ be an element of G. The elements $\{\overline{c\sigma}^{-1}w_j\}_{c,j}$ form a basis for E over k.

We observe that $\bar{c}\sigma\overline{c\sigma}^{-1}$ is an element of S because

$$S\bar{c}\sigma = Sc\sigma = S\overline{c\sigma}.$$

We have

$$\sigma(\overline{c\sigma}^{-1}w_j) = \bar{c}^{-1}(\bar{c}\sigma\overline{c\sigma}^{-1})w_j.$$

Let

$$(\bar{c}\sigma\overline{c\sigma}^{-1})_{\mu j}$$

be the components of the matrix representing the effect of $\bar{c}\sigma\overline{c\sigma}^{-1}$ on F with respect to the basis $\{w_1, \ldots, w_m\}$. Then the action of σ on E is given by

$$\begin{aligned}\sigma(\overline{c\sigma}^{-1}w_j) &= \bar{c}^{-1}\sum_\mu (\bar{c}\sigma\overline{c\sigma}^{-1})_{\mu j}w_\mu \\ &= \sum_\mu (\bar{c}\sigma\overline{c\sigma}^{-1})_{\mu j}(\bar{c}^{-1}w_\mu).\end{aligned}$$

By definition,

$$\chi(\sigma) = \sum_{c\sigma = c}\sum_j (\bar{c}\sigma\overline{c\sigma}^{-1})_{jj}.$$

But $c\sigma = c$ if and only if $\bar{c}\sigma\bar{c}^{-1} \in S$. Furthermore,

$$\psi(\bar{c}\sigma\bar{c}^{-1}) = \sum_j (\bar{c}\sigma\bar{c}^{-1})_{jj}.$$

Hence

$$\chi(\sigma) = \sum_c \psi_0(\bar{c}\sigma\bar{c}^{-1}),$$

as was to be shown.

Remark. Having given an explicit description of the representation space for an induced character, we have in some sense completed the more elementary part of the theory of induced characters. Readers interested in seeing an application can immediately read §12.

Double cosets

Let G be a group and let S be a subgroup. To avoid superscripts we use the following notation. Let $\gamma \in G$. We write

$$[\gamma]S = \gamma S \gamma^{-1} \quad \text{and} \quad S[\gamma] = \gamma^{-1} S \gamma.$$

We shall suppose that S has finite index. We let H be a subgroup. A subset of G of the form $H\gamma S$ is called a **double coset**. As with cosets, it is immediately verified that G is a disjoint union of double cosets. We let $\{\gamma\}$ be a family of double coset representatives, so we have the disjoint union

$$G = \bigcup_\gamma H\gamma S.$$

For each γ we have a decomposition into ordinary cosets

$$H = \bigcup_{\tau_\gamma} \tau_\gamma (H \cap [\gamma]S),$$

where $\{\tau_\gamma\}$ is a finite family of elements of H, depending on γ.

Lemma 7.5. *The elements $\{\tau_\gamma \gamma\}$ form a family of left coset representatives for S in G; that is, we have a disjoint union*

$$G = \bigcup_{\gamma, \tau_\gamma} \tau_\gamma \gamma S.$$

Proof. First we have by hypothesis

$$G = \bigcup_\gamma \bigcup_{\tau_\gamma} \tau_\gamma (H \cap [\gamma]S)\gamma S,$$

and so every element of G can be written in the form

$$\tau_\gamma \gamma s_1 \gamma^{-1} \gamma s_2 = \tau_\gamma \gamma s \quad \text{with} \quad s_1, s_2, s \in S.$$

On the other hand, the elements $\tau_\gamma \gamma$ represent distinct cosets of S, because if $\tau_\gamma \gamma S = \tau_{\gamma'} \gamma' S$, then $\gamma = \gamma'$, since the elements γ represent distinct double cosets,

whence τ_γ and $\tau_{\gamma'}$ represent the same coset of $\gamma S \gamma^{-1}$, and therefore are equal. This proves the lemma.

Let F be an S-module. Given $\gamma \in G$, we denote by $[\gamma]F$ the $[\gamma]S$-module such that for $\gamma s \gamma^{-1} \in [\gamma]S$, the operation is given by

$$\gamma s \gamma^{-1} \cdot [\gamma]x = [\gamma]sx.$$

This notation is compatible with the notation that if F is a submodule of a G-module E, then we may form γF either according to the formal definition above, or according to the operation of G. The two are naturally isomorphic (essentially equal). We shall write

$$[\gamma] : F \to \gamma F \text{ or } [\gamma]F$$

for the above isomorphism from the S-module F to the $[\gamma]S$-module γF. If S_1 is a subgroup of S, then by restriction F is also an S_1-module, and we use $[\gamma]$ also in this context, especially for the subgroup $H \cap [\gamma]S$ which is contained in $[\gamma]S$.

Theorem 7.6. *Applied to the S-module F, we have an isomorphism of H-modules*

$$\operatorname{res}_H^G \circ \operatorname{ind}_S^G \approx \bigoplus_\gamma \operatorname{ind}_{H\cap[\gamma]S}^H \circ \operatorname{res}_{H\cap[\gamma]S}^{[\gamma]S} \circ [\gamma]$$

where the direct sum is taken over double coset representatives γ.

Proof. The induced module $\operatorname{ind}_S^G(F)$ is simply the direct sum

$$\operatorname{ind}_S^G(F) = \bigoplus_{\gamma, \tau_\gamma} \tau_\gamma \gamma F$$

by Lemma 7.5, which gives us coset representatives of S in G, and Theorem 7.3. On the other hand, for each γ, the module

$$\bigoplus_{\tau_\gamma} \tau_\gamma \gamma F$$

is a representation module for the induced representation from $H \cap [\gamma]S$ on γF to H. Taking the direct sum over γ, we get the right-hand side of the expression in the theorem, and thus prove the theorem.

Remark. The formal relation of Theorem 7.6 is one which occurred in Artin's formalism of induced characters and L-functions; *cf.* the exercises and [La 70], Chapter XII, §3. For applications to the cohomology of groups, see [La 96]. The formalism also emerged in Mackey's work [Ma 51], [Ma 53], which we shall now consider more systematically. The rest of this section is due to Mackey. For more extensive results and applications, see Curtis-Reiner [CuR 81], especially Chapter 1. See also Exercises 15, 16, and 17.

To deal more systematically with conjugations, we make some general functorial remarks. Let E be a G-module. Possibly one may have a commutative ring R such that E is a (G, R)-module. We shall deal systematically with the functors

Hom_G, $E^\vee$, and the tensor product. Let

$$\lambda : E \to \lambda E$$

by a R-isomorphism. Then interpreting elements of G as endomorphisms of E we obtain a group $\lambda G \lambda^{-1}$ operating on λE. We shall also write $[\lambda]G$ instead of $\lambda G \lambda^{-1}$. Let E_1, E_2 be (G, R)-modules. Let $\lambda_1 : E_i \to \lambda_i E_i$ be R-isomorphisms. Then we have a natural R-isomorphism

$$(1) \qquad \lambda_2 \mathrm{Hom}_G(E_1, E_2)\lambda_1^{-1} = \mathrm{Hom}_{\lambda_2 G \lambda_1^{-1}}(\lambda_1 E_1, \lambda_2 E_2),$$

and especially

$$[\lambda]\mathrm{Hom}_G(E, E) = \mathrm{Hom}_{[\lambda]G}(\lambda E, \lambda E).$$

As a special case of the general situation, let H, S be subgroups of G, and let F_1, F_2 be (H, R)- and (S, R)-modules respectively, and let $\sigma, \tau \in G$. *Suppose that* $\sigma^{-1}\tau$ lies in the double coset $D = H\gamma S$. Then we have an R-isomorphism

$$(2) \qquad \mathrm{Hom}_{[\sigma]H \cap [\tau]S}([\sigma]F_1, [\tau]F_2) \approx \mathrm{Hom}_{H \cap [\gamma]S}(F_1, [\gamma]F_2).$$

This is immediate by conjugation, writing $\tau = \sigma h \gamma s$ with $h \in H$, $s \in S$, conjugating first with $[\sigma h]^{-1}$, and then observing that for $s \in S$, and an S-module F, we have $[s]S = S$, and $[s^{-1}]F$ is isomorphic to F. In light of (2), we see that the R-module on the left-hand side depends only on the double coset. Let D be a double coset. We shall use the notation

$$M_D(F_1, F_2) = \mathrm{Hom}_{H \cap [\gamma]S}(F_1, [\gamma]F_2)$$

where γ represents the double coset D. With this notation we have:

Theorem 7.7. *Let H, S be subgroups of finite index in G. Let F_1, F_2 be (H, R) and (S, R)-modules respectively. Then we have an isomorphism of R-modules*

$$\mathrm{Hom}_G(\mathrm{ind}_H^G(F_1), \mathrm{ind}_S^G(F_2)) \approx \bigoplus_D M_D(F_1, F_2),$$

where the direct sum is taken over all double cosets $H\gamma S = D$.

Proof. We have the isomorphisms:

$$\begin{aligned}\mathrm{Hom}_G(\mathrm{ind}_H^G(F_1), \mathrm{ind}_S^G(F_2)) &\approx \mathrm{Hom}_H(F_1, \mathrm{res}_H^G \circ \mathrm{ind}_S^G(F_2)) \\ &\approx \bigoplus_\gamma \mathrm{Hom}_H(F_1, \mathrm{ind}_{H \cap [\gamma]S}^H \circ \mathrm{res}_{H \cap [\gamma]S}^{[\gamma]S} \circ [\gamma]F_2) \\ &\approx \bigoplus_\gamma \mathrm{Hom}_{H \cap [\gamma]S}(F_1, [\gamma]F_2)\end{aligned}$$

by applying the definition of the induced module in the first and third step, and applying Theorem 7.6 in the second step. Each term in the last expression is what we denoted by $M_D(F_1, F_2)$ if γ is a representative for the double coset D. This proves the theorem.

Corollary 7.8. *Let $R = k = \mathbf{C}$. Let S, H be subgroups of the finite group G. Let $D = H\gamma S$ range over the double cosets, with representatives γ. Let χ be an effective character of H and ψ an effective character of S. Then*

$$\langle \operatorname{ind}_H^G((\chi), \operatorname{ind}_S^G(\psi)\rangle_G = \sum_\gamma \langle \chi, [\gamma]\psi\rangle_{H\cap[\gamma]S}.$$

Proof. Immediate from Theorem 5.17(b) and Theorem 7.7, taking dimensions on the left-hand side and on the right-hand side.

Corollary 7.9. (Irreducibility of the induced character). *Let S be a subgroup of the finite group G. Let $R = k = \mathbf{C}$. Let ψ be an effective character of S. Then $\operatorname{ind}_S^G(\psi)$ is irreducible if and only if ψ is irreducible and*

$$\langle \psi, [\gamma]\psi\rangle_{S\cap[\gamma]S} = 0$$

for all $\gamma \in G$, $\gamma \notin S$.

Proof. Immediate from Corollary 7.8 and Theorem 5.17(a). It is of course trivial that if ψ is reducible, then so is the induced character.

Another way to phrase Corollary 7.9 is as follows. Let F, F' be representation spaces for S (over $\mathbf{C}$). We call F, F' **disjoint** if no simple S-space occurs both in F and F'. Then Corollary 7.9 can be reformulated:

Corollary 7.9′. *Let S be a subgroup of the finite group G. Let F be an (S, k)-space (with $k = \mathbf{C}$). Then $\operatorname{ind}_S^G(F)$ is simple if and only if F is simple and for all $\gamma \in G$ and $\gamma \notin S$, the $S \cap [\gamma]S$-modules F and $[\gamma]F$ are disjoint.*

Next we have the commutation of the dual and induced representations.

Theorem 7.10. *Let S be a subgroup of G and let F be a finite free R-module. Then there is a G-isomorphism*

$$\operatorname{ind}_S^G(F^\vee) \approx (\operatorname{ind}_S^G(F))^\vee.$$

Proof. Let $G = \bigcup \lambda_i S$ be a left coset decomposition. Then, as in Theorem 7.3, we can express the representation space for $\operatorname{ind}_S^G(F)$ as

$$\operatorname{ind}_S^G(F) = \bigoplus \lambda_i F.$$

We may select $\lambda_1 = 1$ (unit element of G). There is a unique R-homomorphism

$$f : F^\vee \to (\operatorname{ind}_S^G(F))^\vee$$

such that for $\varphi \in F^\vee$ and $x \in F$ we have

$$f(\varphi)(\lambda_i x) = \begin{cases} 0 & \text{if } i \neq 1 \\ \varphi(x) & \text{if } i = 1, \end{cases}$$

which is in fact an R-isomorphism of $F^\vee$ on $(\lambda_1 F)^\vee$. We claim that it is an S-

homomorphism. This is a routine verification, which we write down. We have

$$f([\sigma]\varphi)(\lambda_i x) = \begin{cases} 0 & \text{if } i \neq 1 \\ \sigma(\varphi(\sigma^{-1}x)) & \text{if } i = 1. \end{cases}$$

On the other hand, note that if $\sigma \in S$ then $\sigma^{-1}\lambda_1 \in S$ so $\sigma^{-1}\lambda_1 x \in \lambda_1 F$ for $x \in F$; but if $\sigma \notin S$, then $\sigma^{-1}\lambda_i \notin S$ for $i \neq 1$ so $\sigma^{-1}\lambda_i x \notin \lambda_1 F$. Hence

$$[\sigma](f(\varphi))(\lambda_1 x) = \sigma f(\varphi)(\sigma^{-1}\lambda_i x) = \begin{cases} 0 & \text{if } i \neq 1 \\ \sigma(\varphi(\sigma^{-1}x)) & \text{if } i = 1. \end{cases}$$

This proves that f commutes with the action of S.

By the universal property of the induced module, it follows that there is a unique (G, R)-homomorphism

$$\text{ind}_S^G(f) : \text{ind}_S^G(F^\vee) \to (\text{ind}_S^G(F))^\vee,$$

which must be an isomorphism because f was an isomorphism on its image, the λ_1-component of the induced module. This concludes the proof of the theorem.

Theorems and definitions with Hom have analogues with the tensor product. We start with the analogue of the definition.

Theorem 7.11. *Let S be a subgroup of finite index in G. Let F be an S-module, and E a G-module (over the commutative ring R). Then there is an isomorphism*

$$\text{ind}_S^G(\text{res}_S(E) \otimes F) \approx E \otimes \text{ind}_S^G(F).$$

Proof. The G-module $\text{ind}_S^G(F)$ contains F as a summand, because it is the direct sum $\bigoplus \lambda_i F$ with left coset representatives λ_i as in Theorem 7.3. Hence we have a natural S-isomorphism

$$f : \text{res}_S(E) \otimes F \xrightarrow{\approx} E \otimes \lambda_1 F \subset E \otimes \text{ind}_S^G(F).$$

taking the representative λ_1 to be 1 (the unit element of G). By the universal property of induction, there is a G-homomorphism

$$\text{ind}_S^G(f) : \text{ind}_S^G(\text{res}_S(E) \otimes F) \to E \otimes \text{ind}_S^G(F),$$

which is immediately verified to be an isomorphism, as desired. (Note that here it only needed to verify the bijectivity in this last step, which comes from the structure of direct sum as R-modules.)

Before going further, we make some remarks on functorialities. Suppose we have an isomorphism $G \approx G'$, a subgroup H of G corresponding to a subgroup H' of G' under the isomorphism, and an isomorphism $F \approx F'$ from an H-module F to an H'-module F' commuting with the actions of H, H'. Then we get an isomorphism

$$\text{ind}_H^G(F) \approx \text{ind}_{H'}^{G'}(F').$$

In particular, we could take $\sigma \in G$, let $G' = [\sigma]G = G$, $H' = [\sigma]H$ and $F' = [\sigma]F$.

Next we deal with the analogue of Theorem 7.7. We keep the same notation as in that theorem and the discussion preceding it. With the two subgroups H and S, we may then form the tensor product

$$[\sigma]F_1 \otimes [\tau]F_2$$

with $\sigma, \tau \in G$. Suppose $\sigma^{-1}\tau \in D$ for some double coset $D = H\gamma S$. Note that $[\sigma]F_1 \otimes [\tau]F_2$ is a $[\sigma]H \cap [\tau]S$-module. By conjugation we have an isomorphism

$$\text{(3)} \qquad \operatorname{ind}^G_{[\sigma]H\cap[\tau]S}([\sigma]F_1 \otimes [\tau]F_2) \approx \operatorname{ind}^G_{H\cap[\gamma]S}(F_1 \otimes [\gamma]F_2).$$

Theorem 7.12. *There is a G-isomorphism*

$$\operatorname{ind}^G_H(F_1) \otimes \operatorname{ind}^G_S(F_2) \approx \bigoplus_\gamma \operatorname{ind}^G_{H\cap[\gamma]S}(F_1 \otimes [\gamma]F_2),$$

where the sum is taken over double coset representatives γ.

Proof. We have:

$$\operatorname{ind}^G_H(F_1) \otimes \operatorname{ind}^G_S(F_2) \approx \operatorname{ind}^G_H(F_1 \otimes \operatorname{res}_H \operatorname{ind}^G_S(F_2)) \qquad \text{by Theorem 7.11}$$

$$\approx \bigoplus_\gamma \operatorname{ind}^G_H(F_1 \otimes \operatorname{ind}^H_{H\cap[\gamma]S} \operatorname{res}^{[\gamma]S}_{H\cap[\gamma]S}([\gamma]F_2) \qquad \text{by Theorem 7.6}$$

$$\approx \bigoplus_\gamma \operatorname{ind}^G_H\left(\operatorname{ind}^H_{H\cap[\gamma]S}\left(\operatorname{res}^H_{H\cap[\gamma]S}(F_1) \otimes \operatorname{res}^{[\gamma]S}_{H\cap[\gamma]S}([\gamma]F_2)\right)\right) \qquad \text{by Theorem 7.7}$$

$$\approx \bigoplus_\gamma \operatorname{ind}^G_{H\cap[\gamma]S}(F_1 \otimes [\gamma]F_2) \qquad \text{by transitivity of induction}$$

where we view $F_1 \cap [\gamma]F_2$ as an $H \cap [\gamma]S$-module in this last line. This proves the theorem.

General comment. This section has given a lot of relations for the induced representations. In light of the cohomology of groups, each formula may be viewed as giving an isomorphism of functors in dimension 0, and therefore gives rise to corresponding isomorphisms for the higher cohomology groups H^q. The reader may see this developed further than the exercises in [La 96].

Bibliography

[CuR 81] C. W. Curtis and I. Reiner, *Methods of Representation Theory*, John Wiley and Sons, 1981

[La 96] S. Lang, *Topics in cohomology of groups*, Springer Lecture Notes 1996

[La 70] S. Lang, *Algebraic Number Theory*, Addison-Wesley, 1970, reprinted by Springer Verlag, 1986

[Ma 51] G. Mackey, On induced representations of groups, *Amer. J. Math.* **73** (1951), pp. 576–592

[Ma 53] G. Mackey, Symmetric and anti-symmetric Kronecker squares of induced representations of finite groups, *Amer. J. Math.* **75** (1953), pp. 387–405

The next three sections, which are essentially independent of each other, give examples of induced representations. In each case, we show that certain representations are either induced from certain well-known types, or are linear combinations with integral coefficients of certain well-known types. The most striking feature is that we obtain all characters as linear combinations of induced characters arising from 1-dimensional characters. Thus the theory of characters is to a large extent reduced to the study of 1-dimensional, or abelian characters.

§8. POSITIVE DECOMPOSITION OF THE REGULAR CHARACTER

Let G be a finite group and let k be the complex numbers. We let 1_G be the trivial character, and r_G denote the regular character.

Proposition 8.1. *Let H be a subgroup of G, and let ψ be a character of H. Let ψ^G be the induced character. Then the multiplicity of 1_H in ψ is the same as the multiplicity of 1_G in ψ^G.*

Proof. By Theorem 6.1 (i), we have

$$\langle \psi, 1_H \rangle_H = \langle \psi^G, 1_G \rangle_G.$$

These scalar products are precisely the multiplicities in question.

Proposition 8.2. *The regular representation is the representation induced by the trivial character on the trivial subgroup of G.*

Proof. This follows at once from the definition of the induced character

$$\psi^G(\tau) = \sum_{\sigma \in G} \psi_H(\sigma\tau\sigma^{-1}),$$

taking $\psi = 1$ on the trivial subgroup.

Corollary 8.3. *The multiplicity of 1_G in the regular character r_G is equal to 1.*

We shall now investigate the character

$$u_G = r_G - 1_G.$$

Theorem 8.4. (Aramata). *The character nu_G is a linear combination with positive integer coefficients of characters induced by 1-dimensional characters of cyclic subgroups of G.*

The proof consists of two propositions, which give an explicit description of the induced characters. I am indebted to Serre for the exposition, derived from Brauer's.

If A is a cyclic group of order a, we define the function θ_A on A by the conditions:

$$\theta_A(\sigma) = \begin{cases} a & \text{if } \sigma \text{ is a generator of } A \\ 0 & \text{otherwise.} \end{cases}$$

We let $\lambda_A = \varphi(a)r_A - \theta_A$ (where φ is the Euler function), and $\lambda_A = 0$ if $a = 1$.
The desired result is contained in the following two propositions.

Proposition 8.5. *Let G be a finite group of order n. Then*

$$nu_G = \sum \lambda_A^G,$$

the sum being taken over all cyclic subgroups of G.

Proof. Given two class functions χ, ψ on G, we have the usual scalar product:

$$\langle \psi, \chi \rangle_G = \frac{1}{n} \sum_{\sigma \in G} \psi(\sigma)\overline{\chi(\sigma)}.$$

Let ψ be any class function on G. Then:

$$\begin{aligned} \langle \psi, nu_G \rangle &= \langle \psi, nr_G \rangle - \langle \psi, n1_G \rangle \\ &= n\psi(1) - \sum_{\sigma \in G} \psi(\sigma). \end{aligned}$$

On the other hand, using the fact that the induced character is the transpose of the restriction, we obtain

$$\begin{aligned} \sum_A \langle \psi, \lambda_A^G \rangle &= \sum_A \langle \psi | A, \lambda_A \rangle \\ &= \sum_A \langle \psi | A, \varphi(a)r_A - \theta_A \rangle \\ &= \sum_A \varphi(a)\psi(1) - \sum_A \frac{1}{a} \sum_{\sigma \text{ gen } A} a\psi(\sigma) \\ &= n\psi(1) - \sum_{\sigma \in G} \psi(\sigma). \end{aligned}$$

Since the functions on the right and left of the equality sign in the statement of our proposition have the same scalar product with an arbitrary function, they are equal. This proves our proposition.

Proposition 8.6. *If $A \neq \{1\}$, the function λ_A is a linear combination of irreducible nontrivial characters of A with positive integral coefficients.*

Proof. If A is cyclic of prime order, then by Proposition 8.5, we know that $\lambda_A = nu_A$, and our assertion follows from the standard structure of the regular representation.

In order to prove the assertion in general, it suffices to prove that the Fourier coefficients of λ_A with respect to a character of degree 1 are integers $\geqq 0$. Let ψ be a character of degree 1. We take the scalar product with respect to A, and obtain:

$$\begin{aligned}\langle \psi, \lambda_A \rangle &= \varphi(a)\psi(1) - \sum_{\sigma \text{ gen}} \psi(\sigma) \\ &= \varphi(a) - \sum_{\sigma \text{ gen}} \psi(\sigma) \\ &= \sum_{\sigma \text{ gen}} (1 - \psi(\sigma)).\end{aligned}$$

The sum $\sum \psi(\sigma)$ taken over generators of A is an algebraic integer, and is in fact a rational number (for any number of elementary reasons), hence a rational integer. Furthermore, if ψ is non-trivial, all real parts of

$$1 - \psi(\sigma)$$

are > 0 if $\sigma \neq \text{id}$ and are 0 if $\sigma = \text{id}$. From the last two inequalities, we conclude that the sums must be equal to a positive integer. If ψ is the trivial character, then the sum is clearly 0. Our proposition is proved.

Remark. Theorem 8.4 and Proposition 8.6 arose in the context of zeta functions and L-functions, in Aramata's proof that the zeta function of a number field divides the zeta function of a finite extension [Ar 31], [Ar 33]. See also Brauer [Br 47a], [Br 47b]. These results were also used by Brauer in showing an asymptotic behavior in algebraic number theory, namely

$$\log(hR) \sim \log \mathbf{D}^{1/2} \text{ for } [k : \mathbf{Q}]/\log \mathbf{D} \to 0,$$

where h is the number of ideal classes in a number field k, R is the regulator, and $\mathbf{D}$ is the absolute value of the discriminant. For an exposition of this application, see [La 70], Chapter XVI.

Bibliography

[Ar 31] H. Aramata, Über die Teilbarkeit der Dedekindschen Zetafunktionen, *Proc. Imp. Acad. Tokyo* **7** (1931), pp. 334–336

[Ar 33] H. Aramata, Über die Teilbarkeit der Dedekindschen Zetafunktionen, *Proc. Imp. Acad. Tokyo* **9** (1933), pp. 31–34

[Br 47a] R. Brauer, On the zeta functions of algebraic number fields, *Amer. J. Math.* **69** (1947), pp. 243–250

[Br 47b] R. Brauer, On Artin's L-series with general group characters, *Ann. Math.* **48** (1947), pp. 502–514

[La 70] S. Lang, *Algebraic Number Theory*, Springer Verlag (reprinted from Addison-Wesley, 1970)

§9. SUPERSOLVABLE GROUPS

Let G be a finite group. We shall say that G is **supersolvable** if there exists a sequence of subgroups

$$\{1\} \subset G_1 \subset G_2 \subset \cdots \subset G_m = G$$

such that each G_i is normal in G, and G_{i+1}/G_i is cyclic of prime order.

From the theory of p-groups, we know that every p-group is super-solvable, and so is the direct product of a p-group with an abelian group.

Proposition 9.1. *Every subgroup and every factor group of a super-solvable group is supersolvable.*

Proof. Obvious, using the standard homomorphism theorems.

Proposition 9.2. *Let G be a non-abelian supersolvable group. Then there exists a normal abelian subgroup which contains the center properly.*

Proof. Let C be the center of G, and let $\bar{G} = G/C$. Let $\bar{H}$ be a normal subgroup of prime order in $\bar{G}$ and let H be its inverse image in G under the canonical map $G \to G/C$. If σ is a generator of $\bar{H}$, then an inverse image σ of $\bar{\sigma}$, together with C, generate H. Hence H is abelian, normal, and contains the center properly.

Theorem 9.3. (Blichfeldt). *Let G be a supersolvable group, let k be algebraically closed. Let E be a simple (G, k)-space. If* $\dim_k E > 1$, *then there exists a proper subgroup H of G and a simple H-space F such that E is induced by F.*

Proof. Since a simple representation of an abelian group is 1-dimensional, our hypothesis implies that G is not abelian.

We shall first give the proof of our theorem under the additional hypothesis that E is faithful. (This means that $\sigma x = x$ for all $x \in E$ implies $\sigma = 1$.) It will be easy to remove this restriction at the end.

Lemma 9.4. *Let G be a finite group, and assume k algebraically closed. Let E be a simple, faithful G-space over k. Assume that there exists a normal abelian subgroup H of G containing the center of G properly. Then there exists a proper subgroup H_1 of G containing H, and a simple H_1-space F such that E is the induced module of F from H_1 to G.*

Proof. We view E as an H-space. It is a direct sum of simple H-spaces, and since H is abelian, such simple H-space is 1-dimensional.

Let $v \in E$ generate a 1-dimensional H-space. Let ψ be its character. If $w \in E$ also generates a 1-dimensional H-space, with the same character ψ, then

for all $a, b \in k$ and $\tau \in H$ we have

$$\tau(av + bw) = \psi(\tau)(av + bw).$$

If we denote by F_ψ the subspace of E generated by all 1-dimensional H-subspaces having the character ψ, then we have an H-direct sum decomposition

$$E = \bigoplus_\psi F_\psi.$$

We contend that $E \neq F_\psi$. Otherwise, let $v \in E$, $v \neq 0$, and $\sigma \in G$. Then $\sigma^{-1}v$ is a 1-dimensional H-space by assumption, and has character ψ. Hence for $\tau \in H$,

$$\tau(\sigma^{-1}v) = \psi(\tau)\sigma^{-1}v$$

$$(\sigma\tau\sigma^{-1})v = \sigma\psi(\tau)\sigma^{-1}v = \psi(\tau)v.$$

This shows that $\sigma\tau\sigma^{-1}$ and τ have the same effect on the element v of E. Since H is not contained in the center of G, there exist $\tau \in H$ and $\sigma \in G$ such that $\sigma\tau\sigma^{-1} \neq \tau$, and we have contradicted the assumption that E is faithful.

We shall prove that G permutes the spaces F_ψ *transitively.*

Let $v \in F_\psi$. For any $\tau \in H$ and $\sigma \in G$, we have

$$\tau(\sigma v) = \sigma(\sigma^{-1}\tau\sigma)v = \sigma\psi(\sigma^{-1}\tau\sigma)v = \psi_\sigma(\tau)\sigma v,$$

where ψ_σ is the function on H given by $\psi_\sigma(\tau) = \psi(\sigma^{-1}\tau\sigma)$. This shows that σ maps F_ψ into F_{ψ_σ}. However, by symmetry, we see that σ^{-1} maps F_{ψ_σ} into F_ψ, and the two maps σ, σ^{-1} give inverse mappings between F_{ψ_σ} and F_ψ. Thus G permutes the spaces $\{F_\psi\}$.

Let $E' = GF_{\psi_0} = \sum \sigma F_{\psi_0}$ for some fixed ψ_0. Then E' is a G-subspace of E, and since E was assumed to be simple, it follows that $E' = E$. This proves that the spaces $\{F_\psi\}$ are permuted transitively.

Let $F = F_{\psi_1}$ for some fixed ψ_1. Then F is an H-subspace of E. Let H_1 be the subgroup of all elements $\tau \in G$ such that $\tau F = F$. Then $H_1 \neq G$ since $E \neq F_\psi$. *We contend that* F *is a simple* H_1*-subspace, and that* E *is the induced space of* F *from* H_1 *to* G.

To see this, let $G = \bigcup H_1\bar{c}$ be a decomposition of G in terms of right cosets of H_1. Then the elements $\{\bar{c}^{-1}\}$ form a system of left coset representatives of H_1. Since

$$E = \sum_{\sigma \in G} \sigma F$$

it follows that

$$E = \sum_c \bar{c}^{-1}F.$$

We contend that this last sum is direct, and that F is a simple H_1-space.

Since G permutes the spaces $\{F_\psi\}$, we see by definition that H_1 is the isotropy group of F for the operation of G on this set of spaces, and hence that the elements of the orbit are precisely $\{\bar{c}^{-1}F\}$, as c ranges over all the cosets. Thus the spaces $\{\bar{c}^{-1}F\}$ are distinct, and we have a direct sum decomposition

$$E = \bigoplus_c \bar{c}^{-1}F.$$

If W is a proper H_1-subspace of F, then $\bigoplus \bar{c}^{-1}W$ is a proper G-subspace of E, contradicting the hypothesis that E is simple. This proves our assertions.

We can now apply Theorem 7.3 to conclude that E is the induced module from F, thereby proving Theorem 9.3, in case E is assumed to be faithful.

Suppose now that E is not faithful. Let G_0 be the normal subgroup of G which is the kernel of the representation $G \to \mathrm{Aut}_k(E)$. Let $\bar{G} = G/G_0$. Then E gives a faithful representation of $\bar{G}$. As E is not 1-dimensional, then $\bar{G}$ is not abelian and there exists a proper normal subgroup $\bar{H}$ of $\bar{G}$ and a simple $\bar{H}$-space F such that

$$E = \mathrm{ind}_{\bar{H}}^{\bar{G}}(F).$$

Let H be the inverse image of $\bar{H}$ in the natural map $G \to \bar{G}$. Then $H \supset G_0$, and F is a simple H-space. In the operation of $\bar{G}$ as a permutation group of the k-subspaces $\{\sigma F\}_{\sigma \in G}$, we know that $\bar{H}$ is the isotropy group of one component. Hence H is the isotropy group in G of this same operation, and hence applying Theorem 7.3 again, we conclude that E is induced by F in G, i.e.

$$E = \mathrm{ind}_H^G(F),$$

thereby proving Theorem 9.3.

Corollary 9.5. *Let G be a product of a p-group and a cyclic group, and let k be algebraically closed. If E is a simple (G, k)-space and is not 1-dimensional, then E is induced by a 1-dimensional representation of some subgroup.*

Proof. We apply the theorem step by step using the transitivity of induced representations until we get a 1-dimensional representation of a subgroup.

§10. BRAUER'S THEOREM

We let $k = \mathbf{C}$ be the field of complex numbers. We let R be a subring of k. We shall deal with $X_R(G)$, i.e. the ring consisting of all linear combinations with coefficients in R of the simple characters of G over k. (It is a ring by Proposition 2.1.)

Let $H = \{H_\alpha\}$ be a fixed family of subgroups of G, indexed by indices $\{\alpha\}$. We let $V_R(G)$ be the additive subgroup of $X_R(G)$ generated by all the functions which are induced by functions in $X_R(H_\alpha)$ for some H_α in our family. In other words,

$$V_R(G) = \sum_\alpha \operatorname{ind}_{H_\alpha}^G(X_R(H_\alpha)).$$

We could also say that $V_R(G)$ is the subgroup generated over R by all the characters induced from all the H_α.

Lemma 10.1. *$V_R(G)$ is an ideal in $X_R(G)$.*

Proof. This is immediate from Theorem 6.1.

For many applications, the family of subgroups will consist of "elementary" subgroups: Let p be a prime number. By a ***p*-elementary group** we shall mean the product of a p-group and a cyclic group (whose order may be assumed prime to p, since we can absorb the p-part of a cyclic factor into the p-group). An element $\sigma \in G$ is said to be ***p*-regular** if its period is prime to p, and ***p*-singular** if its period is a power of p. Given $x \in G$, we can write in a unique way

$$x = \sigma\tau$$

where σ is p-singular, τ is p-regular, and σ, τ commute. Indeed, if $p^r m$ is the period of x, with m prime to p, then $1 = \nu p^r + \mu m$ whence $x = (x^m)^\mu (x^{p^r})^\nu$ and we get our factorization. It is clearly unique, since the factors have to lie in the cyclic subgroup generated by x. We call the two factors the ***p*-singular** and ***p*-regular factors** of x respectively.

The above decomposition also shows:

Proposition 10.2. *Every subgroup and every factor group of a p-elementary group is p-elementary. If S is a subgroup of the p-elementary group $P \times C$, where P is a p-group, and C is cyclic, of order prime to p, then*

$$S = (S \cap P) \times (S \cap C).$$

Proof. Clear.

Our purpose is to show, among other things, that if our family $\{H_\alpha\}$ is such that every p-elementary subgroup of G is contained in some H_α, then $V_R(G) = X_R(G)$ for every ring R. It would of course suffice to do it for $R = \mathbf{Z}$, but for our purposes, it is necessary to prove the result first using a bigger ring. The main result is contained in Theorems 10.11 and 10.13, due to Brauer. We shall give an exposition of Brauer-Tate (*Annals of Math.*, July 1955).

We let R be the ring $\mathbf{Z}[\zeta]$ where ζ is a primitive n-th root of unity. There exists a basis of R as a $\mathbf{Z}$-module, namely $1, \zeta, \dots, \zeta^{N-1}$ for some integer N. This is a trivial fact, and we can take N to be the degree of the irreducible polynomial of ζ over $\mathbf{Q}$. This irreducible polynomial has leading coefficient 1, and

has integer coefficients, so the fact that

$$1, \zeta, \ldots, \zeta^{N-1}$$

form a basis of $\mathbf{Z}[\zeta]$ follows from the Euclidean algorithm. We don't need to know anything more about this degree N.

We shall prove our assertion first for the above ring R. The rest then follows by using the following lemma.

Lemma 10.3. *If $d \in \mathbf{Z}$ and the constant function $d.1_G$ belongs to V_R then $d.1_G$ belongs to $V_{\mathbf{Z}}$.*

Proof. We contend that $1, \zeta, \ldots, \zeta^{N-1}$ are linearly independent over $X_{\mathbf{Z}}(G)$. Indeed, a relation of linear dependence would yield

$$\sum_{\nu=1}^{s} \sum_{j=0}^{N-1} c_{\nu j} \chi_\nu \zeta^j = 0$$

with integers $c_{\nu j}$ not all 0. But the simple characters are linearly independent over k. The above relation is a relation between these simple characters with coefficients in R, and we get a contradiction. We conclude therefore that

$$V_R = V_{\mathbf{Z}} \oplus V_{\mathbf{Z}}\zeta \oplus \cdots \oplus V_{\mathbf{Z}}\zeta^{N-1}$$

is a direct sum (of abelian groups), and our lemma follows.

If we can succeed in proving that the constant function 1_G lies in $V_R(G)$, then by the lemma, we conclude that it lies in $V_{\mathbf{Z}}(G)$, and since $V_{\mathbf{Z}}(G)$ is an ideal, that $X_{\mathbf{Z}}(G) = V_{\mathbf{Z}}(G)$.

To prove our theorem, we need a sequence of lemmas.

Two elements x, x' of G are said to be ***p*-conjugate** if their p-regular factors are conjugate in the ordinary sense. It is clear that p-conjugacy is an equivalence relation, and an equivalence class will be called a ***p*-conjugacy class**, or simply a ***p*-class**.

Lemma 10.4. *Let $f \in X_R(G)$, and assume that $f(\sigma) \in \mathbf{Z}$ for all $\sigma \in G$. Then f is constant* mod p *on every p-class.*

Proof. Let $x = \sigma\tau$, where σ is p-singular, and τ is p-regular, and σ, τ commute. It will suffice to prove that

$$f(x) \equiv f(\tau) \pmod{p}.$$

Let H be the cyclic subgroup generated by x. Then the restriction of f to H can be written

$$f_H = \sum a_j \psi_j$$

with $a_j \in R$, and ψ_j being the simple characters of H, hence homomorphisms of H into k^*. For some power p^r we have $x^{p^r} = \tau^{p^r}$, whence $\psi_j(x)^{p^r} = \psi_j(\tau)^{p^r}$, and hence

$$f(x)^{p^r} \equiv f(\tau)^{p^r} \pmod{pR}.$$

We now use the following lemma.

Lemma 10.5. *Let $R = \mathbf{Z}[\zeta]$ be as before. If $a \in \mathbf{Z}$ and $a \in pR$ then $a \in p\mathbf{Z}$.*

Proof. This is immediate from the fact that R has a basis over $\mathbf{Z}$ such that 1 is a basis element.

Applying Lemma 10.5, we conclude that $f(x) \equiv f(\tau) \pmod{p}$, because $b^{p^r} \equiv b \pmod{p}$ for every integer b.

Lemma 10.6. *Let τ be p-regular in G, and let T be the cyclic subgroup generated by τ. Let C be the subgroup of G consisting of all elements commuting with τ. Let P be a p-Sylow subgroup of C. Then there exists an element $\psi \in X_R(T \times P)$ such that the induced function $f = \psi^G$ has the following properties:*

(i) *$f(\sigma) \in \mathbf{Z}$ for all $\sigma \in G$.*

(ii) *$f(\sigma) = 0$ if σ does not belong to the p-class of τ.*

(iii) *$f(\tau) = (C : P) \not\equiv 0 \pmod{p}$.*

Proof. We note that the subgroup of G generated by T and P is a direct product $T \times P$. Let $\psi_1, \ldots, \psi_r$ be the simple characters of the cyclic group T, and assume that these are extended to $T \times P$ by composition with the projection:

$$T \times P \to T \to k^*.$$

We denote the extensions again by $\psi_1, \ldots, \psi_r$. Then we let

$$\psi = \sum_{\nu=1}^{r} \overline{\psi_\nu(\tau)}\psi_\nu.$$

The orthogonality relations for the simple characters of T show that

$$\psi(\tau y) = \psi(\tau) = (T : 1) \quad \text{for} \quad y \in P$$

$$\psi(\sigma) = 0 \quad \text{if} \quad \sigma \in TP, \quad \text{and} \quad \sigma \notin \tau P.$$

We contend that ψ^G satisfies our requirements.

First, it is clear that ψ lies in $X_R(TP)$.

We have for $\sigma \in G$:

$$\psi^G(\sigma) = \frac{1}{(TP:1)} \sum_{x \in G} \psi_{TP}(x\sigma x^{-1}) = \frac{1}{(P:1)} \mu(\sigma)$$

where $\mu(\sigma)$ is the number of elements $x \in G$ such that $x\sigma x^{-1}$ lies in τP. The number $\mu(\sigma)$ is divisible by $(P:1)$ because if an element x of G moves σ into τP by conjugation, so does every element of Px. Hence the values of ψ^G lie in $\mathbf{Z}$.

Furthermore, $\mu(\sigma) \neq 0$ only if σ is p-conjugate to τ, whence our condition (ii) follows.

Finally, we can have $x\tau x^{-1} = \tau y$ with $y \in P$ only if $y = 1$ (because the period of τ is prime to p). Hence $\mu(\tau) = (C:1)$, and our condition (iii) follows.

Lemma 10.7. *Assume that the family of subgroups $\{H_\alpha\}$ covers G (i.e. every element of G lies in some H_α). If f is a class function on G taking its values in $\mathbf{Z}$, and such that all the values are divisible by $n = (G:1)$, then f belongs to $V_R(G)$.*

Proof. Let γ be a conjugacy class, and let p be prime to n. Every element of G is p-regular, and all p-subgroups of G are trivial. Furthermore, p-conjugacy is the same as conjugacy. Applying Lemma 10.6, we find that there exists in $V_R(G)$ a function taking the value 0 on elements $\sigma \notin \gamma$, and taking an integral value dividing n on elements of γ. Multiplying this function by some integer, we find that there exists a function in $V_R(G)$ taking the value n for all elements of γ, and the value 0 otherwise. The lemma then follows immediately.

Theorem 10.8. (Artin). *Every character of G is a linear combination with rational coefficients of induced characters from cyclic subgroups.*

Proof. In Lemma 10.7, let $\{H_\alpha\}$ be the family of cyclic subgroups of G. The constant function $n.1_G$ belongs to $V_R(G)$. By Lemma 10.3, this function belongs to $V_{\mathbf{Z}}(G)$, and hence $nX_{\mathbf{Z}}(G) \subset V_{\mathbf{Z}}(G)$. Hence

$$X_{\mathbf{Z}}(G) \subset \frac{1}{n} V_{\mathbf{Z}}(G),$$

thereby proving the theorem.

Lemma 10.9. *Let p be a prime number, and assume that every p-elementary subgroup of G is contained in some H_α. Then there exists a function $f \in V_R(G)$ whose values are in $\mathbf{Z}$, and $\equiv 1 \pmod{p^r}$.*

Proof. We apply Lemma 10.6 again. For each p-class γ, we can find a function f_γ in $V_R(G)$, whose values are 0 on elements outside γ, and $\not\equiv 0 \bmod p$ for elements of γ. Let $f = \sum f_\gamma$, the sum being taken over all p-classes. Then $f(\sigma) \not\equiv 0 \pmod{p}$ for all $\sigma \in G$. Taking $f^{(p-1)p^{r-1}}$ gives what we want.

Lemma 10.10. *Let p be a prime number and assume that every p-elementary subgroup of G is contained in some H_α. Let $n = n_0 p^r$ where n_0 is prime to p. Then the constant function $n_0 . 1_G$ belongs to $V_{\mathbf{Z}}(G)$.*

Proof. By Lemma 10.3, it suffices to prove that $n_0 . 1_G$ belongs to $V_R(G)$. Let f be as in Lemma 10.9. Then

$$n_0 . 1_G = n_0(1_G - f) + n_0 f.$$

Since $n_0(1_G - f)$ has values divisible by $n_0 p^r = n$, it lies in $V_R(G)$ by Lemma 10.7. On the other hand, $n_0 f \in V_R(G)$ because $f \in V_R(G)$. This proves our lemma.

Theorem 10.11. (Brauer). *Assume that for every prime number p, every p-elementary subgroup of G is contained in some H_α. Then $X(G) = V_{\mathbf{Z}}(G)$. Every character of G is a linear combination, with integer coefficients, of characters induced from subgroups H_α.*

Proof. Immediate from Lemma 10.10, since we can find functions $n_0 . 1_G$ in $V_{\mathbf{Z}}(G)$ with n_0 relatively prime to any given prime number.

Corollary 10.12. *A class function f on G belongs to $X(G)$ if and only if its restriction to H_α belongs to $X(H_\alpha)$ for each α.*

Proof. Assume that the restriction of f to H_α is a character on H_α for each α. By the theorem, we can write

$$1_G = \sum_\alpha c_\alpha \operatorname{ind}_{H_\alpha}^G(\psi_\alpha)$$

where $c_\alpha \in \mathbf{Z}$, and $\psi_\alpha \in X(H_\alpha)$. Hence

$$f = \sum_\alpha c_\alpha \operatorname{ind}_{H_\alpha}^G(\psi_\alpha f_{H_\alpha}),$$

using Theorem 6.1. If $f_{H_\alpha} \in X(H_\alpha)$, we conclude that f belongs to $X(G)$. The converse is of course trivial.

Theorem 10.13. (Brauer). *Every character of G is a linear combination with integer coefficients of characters induced by 1-dimensional characters of subgroups.*

Proof. By Theorem 10.11, and the transitivity of induction, it suffices to prove that every character of a p-elementary group has the property stated in the theorem. But we have proved this in the preceding section, Corollary 9.5.

§11. FIELD OF DEFINITION OF A REPRESENTATION

We go back to the general case of k having characteristic prime to $\#G$. Let E be a k-space and assume we have a representation of G on E. Let k' be an extension field of k. Then G operates on $k' \otimes_k E$ by the rule

$$\sigma(a \otimes x) = a \otimes \sigma x$$

for $a \in k'$ and $x \in E$. This is obtained from the bilinear map on the product $k' \times E$ given by

$$(a, x) \mapsto a \otimes \sigma x.$$

We view $E' = k' \otimes_k E$ as the extension of E by k', and we obtain a representation of G on E'.

Proposition 11.1. *Let the notation be as above. Then the characters of the representations of G on E and on E' are equal.*

Proof. Let $\{v_1, \ldots, v_m\}$ be a basis of E over k. Then

$$\{1 \otimes v_1, \ldots, 1 \otimes v_m\}$$

is a basis of E' over k'. Thus the matrices representing an element σ of G with respect to the two bases are equal, and consequently the traces are equal.

Conversely, let k' be a field and k a subfield. A representation of G on a k'-space E' is said to be **definable over** k if there exists a k-space E and a representation of G on E such that E' is G-isomorphic to $k' \otimes_k E$.

Proposition 11.2. *Let E, F be simple representation spaces for the finite group G over k. Let k' be an extension of k. Assume that E, F are not G-isomorphic. Then no k'-simple component of $E_{k'}$ appears in the direct sum decomposition of $F_{k'}$ into k'-simple subspaces.*

Proof. Consider the direct product decomposition

$$k[G] = \prod_{\mu=1}^{s(k)} R_\mu(k)$$

over k, into a direct product of simple rings. Without loss of generality, we may assume that E, F are simle left ideals of $k[G]$, and they will belong to distinct factors of this product by assumption. We now take the tensor product with k', getting nothing else but $k'[G]$. Then we obtain a direct product decomposition over k'. Since $R_\nu(k)R_\mu(k) = 0$ if $\nu \neq \mu$, this will actually be given by a direct

product decomposition of each factor $R_\mu(k)$:

$$k'[G] = \prod_{\mu=1}^{s(k)} \prod_{i=1}^{m(\mu)} R_{\mu i}(k').$$

Say $E = L_\nu$ and $F = L_\mu$ with $\nu \neq \mu$. Then $R_\mu E = 0$. Hence $R_{\mu i} E_{k'} = 0$ for each $i = 1, \ldots, m(\mu)$. This implies that no simple component of $E_{k'}$ can be G-isomorphic to any one of the simple left ideals of $R_{\mu i}$, and proves what we wanted.

Corollary 11.3. *The simple characters $\chi_1, \ldots, \chi_{s(k)}$ of G over k are linearly independent over any extension k' of k.*

Proof. This follows at once from the proposition, together with the linear independence of the k'-simple characters over k'.

Propositions 11.1 and 11.2 are essentially general statements of an abstract nature. The next theorem uses Brauer's theorem in its proof.

Theorem 11.4. (Brauer). *Let G be a finite group of exponent m. Every representation of G over the complex numbers (or an algebraically closed field of characteristic 0) is definable over the field $\mathbf{Q}(\zeta_m)$ where ζ_m is a primitive m-th root of unity.*

Proof. Let χ be the character of a representation of G over $\mathbf{C}$, i.e. an effective character. By Theorem 10.13, we can write

$$\chi = \sum_j c_j \operatorname{ind}_{S_j}^G(\psi_j), \qquad c_j \in \mathbf{Z},$$

the sum being taken over a finite number of subgroups S_j, and ψ_j being a 1-dimensional character of S_j. It is clear that each ψ_j is definable over $\mathbf{Q}(\zeta_m)$. Thus the induced character ψ_j^G is definable over $\mathbf{Q}(\zeta_m)$. Each ψ_j^G can be written

$$\psi_j^G = \sum_\mu d_{j\mu} \chi_\mu, \qquad d_{j\mu} \in \mathbf{Z}$$

where $\{\chi_\mu\}$ are the simple characters of G over $\mathbf{Q}(\zeta_m)$. Hence

$$\chi = \sum_\mu \left(\sum_j c_j d_{j\mu} \right) \chi_\mu .$$

The expression of χ as a linear combination of the simple characters over k is unique, and hence the coefficient

$$\sum_j c_j d_{j\mu}$$

is $\geqq 0$. This proves what we wanted.

§12. EXAMPLE: GL_2 OVER A FINITE FIELD

Let F be a field. We view $GL_2(F)$ as operating on the 2-dimensional vector space $V = F^2$. We let F^a be the algebraic closure as usual, and we let $V^a = F^a \times F^a = F^a \otimes V$ (tensor product over F). By **semisimple**, we always mean absolutely semisimple, i.e. semisimple over the algebraic closure F^a. An element $\alpha \in GL_2(F)$ is called **semisimple** if V^a is semisimple over $F^a[\alpha]$. A subgroup is called **semisimple** if all its elements are semisimple.

Let K be a separable quadratic extension of F. Let $\{\omega_1, \omega_2\}$ be a basis of K. Then we have the regular representation of K with respect to this basis, namely multiplication representing K^* as a subgroup of $GL_2(F)$. The elements of norm 1 correspond precisely to the elements of $SL_2(F)$ in the image of K^*. A different choice of basis of K corresponds to conjugation of this image in $GL_2(F)$. Let C_K denote one of these images. Then C_K is called a **non-split Cartan subgroup**. The subalgebra

$$F[C_K] \subset \mathrm{Mat}_2(F)$$

is isomorphic to K itself, and the units of the algebra are therefore the elements of $C_K \approx K^*$.

Lemma 12.1. *The subgroup C_K is a maximal commutative semisimple subgroup.*

Proof. If $\alpha \in GL_2(F)$ commutes with all elements of C_K then α must lie in $F[C_K]$, for otherwise $\{1, \alpha\}$ would be linearly independent over $F[C_K]$, whence $\mathrm{Mat}_2(F)$ would be commutative, which is not the case. Since α is invertible, α is a unit in $F[C_K]$, so $\alpha \in C_K$, as was to be shown.

By the **split Cartan subgroup** we mean the group of diagonal matrices

$$\begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix} \text{ with } a, d \in F^*.$$

We denote the split Cartan by A, or $A(F)$ if the reference to F is needed.

By a **Cartan subgroup** we mean a subgroup conjugate to the split Cartan or to one of the subgroups C_K as above.

Lemma 12.2. *Every maximal commutative semisimple subgroup of $GL_2(F)$ is a Cartan subgroup, and conversely.*

Proof. It is clear that the split Cartan subgroup is maximal commutative semisimple. Suppose that H is a maximal commutative semisimple subgroup of $GL_2(F)$. If H is diagonalizable over F, then H is contained in a conjugate of the split Cartan. On the other hand, suppose H is not diagonalizable over F. It is diagonalizable over the separable closure of F, and the two eigenspaces of

dimension 1 give rise to two characters

$$\psi, \psi' : H \to F^{s*}$$

of H in the multiplicative group of the separable closure. For each element $\alpha \in H$ the values $\psi(\alpha)$ and $\psi'(\alpha)$ are the eigenvalues of α, and for some element $\alpha \in H$ these eigenvalues are distinct, otherwise H is diagonalizable over F. Hence the pair of elements $\psi(\alpha)$, $\psi'(\alpha)$ are conjugate over F. The image $\psi(H)$ is cyclic, and if $\psi(\alpha)$ generates this image, then we see that $\psi(\alpha)$ generates a quadratic extension K of F. The map

$$\alpha \mapsto \psi(\alpha) \text{ with } \alpha \in H$$

extends to an F-linear mapping, also denoted by ψ, of the algebra $F[H]$ into K. Since $F[H]$ is semisimple, it follows that $\psi : F[H] \to K$ is an isomorphism. Hence ψ maps H into K^*, and in fact maps H onto K^* because H was taken to be maximal. This proves the lemma.

In the above proof, the two characters ψ, ψ' are called the **(eigen)characters of the Cartan subgroup**. In the split case, if α has diagonal elements, a, d then we get the two characters such that $\psi(\alpha) = a$ and $\psi'(\alpha) = d$. In the split case, the values of the characters are in F. In the non-split case, these values are conjugate quadratic over F, and lie in K.

Proposition 12.3. *Let H be a Cartan subgroup of $GL_2(F)$ (split or not). Then H is of index 2 in its normalizer $N(H)$.*

Proof. We may view $GL_2(F)$ as operating on the 2-dimensional vector space $V^a = F^a \oplus F^a$, over the algebraic closure F^a. Whether H is split or not, the eigencharacters are distinct (because of the separability assumption in the non-split case), and an element of the normalizer must either fix or interchange the eigenspaces. If it fixes them, then it lies in H by the maximality of H in Lemma 12.2. If it interchanges them, then it does not lie in H, and generates a unique coset of N/H, so that H is of index 2 in N.

In the split case, a representative of N/A which interchanges the eigenspaces is given by

$$w = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

In the non-split case, let $\sigma: K \to K$ be the non-trivial automorphism. Let $\{\alpha, \sigma\alpha\}$ be a normal basis. With respect to this basis, the matrix of σ is precisely the matrix

$$w = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Therefore again in this case we see that there exists a non-trivial element in the

normalizer of A. Note that it is immediate to verify the relation

$$M(\sigma)M(x)M(\sigma^{-1}) = M(\sigma x),$$

if $M(x)$ is the matrix associated with an element $x \in K$.

Since the order of an element in the multiplicative group of a field is prime to the characteristic, we conclude:

If F has characteristic p, then an element of finite order in $GL_2(F)$ is semisimple if and only if its order is prime to p.

Conjugacy classes

We shall determine the conjugacy classes explicitly. We specialize the situation, and from now on we let:

F = finite field with q elements;
$G = GL_2(F)$;
Z = center of G;
A = diagonal subgroup of G;
$C \approx K^*$ = a non-split Cartan subgroup of G.

Up to conjugacy there is only one non-split Cartan because over a finite field there is only one quadratic extension (in a given algebraic closure F^{a}) (*cf.* Corollary 2.7 of Chapter XIV). Recall that

$$\#(G) = (q^2 - 1)(q^2 - q) = q(q + 1)(q - 1)^2.$$

This should have been worked out as an exercise before. Indeed, $F \times F$ has q^2 elements, and $\#(G)$ is equal to the number of bases of $F \times F$. There are $q^2 - 1$ choices for a first basis element, and then $q^2 - q$ choices for a second (omitting $(0, 0)$ the first time, and all chosen elements the second time). This gives the value for $\#(G)$.

There are two cases for the conjugacy classes of an element α.

Case 1. The characteristic polynomial is reducible, so the eigenvalues lie in F. In this case, by the Jordan canonical form, such an element is conjugate to one of the matrices

$$\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}, \quad \begin{pmatrix} a & 1 \\ 0 & a \end{pmatrix}, \quad \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix} \quad \text{with } d \neq a.$$

These are called **central, unipotent**, or **rational not central** respectively.

Case 2. The characteristic polynomial is irreducible. Then α is such that $F[\alpha] \approx E$, where E is the quadratic extension of F of degree 2. Then $\{1, \alpha\}$ is a basis of $F[\alpha]$ over F, and the matrix associated with α under the representation by multiplication on $F[\alpha]$ is

$$\begin{pmatrix} 0 & -b \\ 1 & -a \end{pmatrix},$$

where a, b are the coefficients of the characteristic polynomial $X^2 + ax + b$.
We then have the following table.

Table 12.4

class	# of classes	# of elements in the class
$\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$	$q - 1$	1
$\begin{pmatrix} a & 1 \\ 0 & a \end{pmatrix}$	$q - 1$	$q^2 - 1$
$\begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}$ with $a \neq d$	$\frac{1}{2}(q - 1)(q - 2)$	$q^2 + q$
$\alpha \in C - F^*$	$\frac{1}{2}(q - 1)q$	$q^2 - q$

In each case one computes the number of elements in a given class as the index of the normalizer of the element (or centralizer of the element). Case 1 is trivial. Case 2 can be done by direct computation, since the centralizer is then seen to consist of the matrices

$$\begin{pmatrix} x & y \\ 0 & x \end{pmatrix}, \; x \in F,$$

with $x \neq 0$. The third and fourth cases can be done by using Proposition 12.3.

As for the number of classes of each type, the first and second cases correspond to distinct choices of $a \in F^*$ so the number of classes is $q - 1$ in each case. In the third case, the conjugacy class is determined by the eigenvalues. There are $q - 1$ possible choices for a, and then $q - 2$ possible choices for d. But the non-ordered pair of eigenvalues determines the conjugacy class, so one must divide $(q-1)(q-2)$ by 2 to get the number of classes. Finally, in the case of an element in a non-split Cartan, we have already seen that if σ generates $\mathrm{Gal}(K/F)$, then $M(\sigma x)$ is conjugate to $M(x)$ in $GL_2(F)$. But on the other hand, suppose $x, x' \in K^*$ and $M(x), M(x')$ are conjugate in $GL_2(F)$ under a given regular representation of K^* on K with respect to a given basis. Then this conjugation induces an F-algebra isomorphism on $F[C_K]$, whence an automorphism of K, which is the identity, or the non-trivial automorphism σ. Consequently the number of conjugacy classes for elements of the fourth type is equal to

$$\frac{\#(K) - \#(F)}{2} = \frac{q^2 - q}{2},$$

which gives the value in the table.

Borel subgroup and induced representations

We let:

U = group of unipotent elements $\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}$;

B = Borel subgroup = $UA = AU$.

Then $\#(B) = q(q-1)^2 = (q-1)(q^2-q)$. We shall construct representations of G by inducing characters from B, and eventually we shall construct all irreducible representations of G by combining the induced representations in a suitable way. We shall deal with four types of characters. Except in the first type, which is 1-dimensional and therefore obviously simple, we shall prove that the other types are simple by computing induced characters. In one case we need to subtract a one-dimensional character. In the other cases, the induced character will turn out to be simple. The procedure will be systematic. We shall give a table of values for each type. We verify in each case that for the character χ which we want to prove simple we have

$$\sum_{\beta \in G} |\chi(\beta)|^2 = \#(G),$$

and then apply Theorem 5.17(a) to get the simplicity. Once we have done this for all four types, from the tables of values we see that they are distinct. Finally, the total number of distinct characters which we have exhibited will be equal to the number of conjugacy classes, whence we conclude that we have exhibited all simple characters.

We now carry out this program. I myself learned the simple characters of $GL_2(F)$ from a one-page handout by Tate in a course at Harvard, giving the subsequent tables and the values of the characters on conjugacy classes. I filled out the proofs in the following pages.

First type

$\mu : F^* \to C^*$ denotes a homomorphism. Then we obtain the character

$\mu \circ \det : G \to \mathbf{C}^*$,

which is 1-dimensional. Its values on representatives of the conjugacy classes are given in the following table.

Table 12.5(I)

χ	$\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$	$\begin{pmatrix} a & 1 \\ 0 & a \end{pmatrix}$	$\begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix} d \neq a$	$\alpha \in C - F^*$
$\mu \circ \det$	$\mu(a)^2$	$\mu(a)^2$	$\mu(ad)$	$\mu \circ \det(\alpha)$

The stated values are by definition. The last value can also be written

$$\mu(\det \alpha) = \mu(N_{K/F}(\alpha)),$$

viewing α as an element of K^*, because the reader should know from field theory that the determinant gives the norm.

A character of G will be said to be of **first type** if it is equal to $\mu \circ \det$ for some μ. There are $q - 1$ characters of first type, because $\#(F^*) = q - 1$.

Second type

Observe that we have $B/U = A$. A character of A can therefore be viewed as a character on B via B/U. We let:

$\psi_\mu = \mathrm{res}_A(\mu \circ \det)$, and view ψ_μ therefore as a character on B. Thus

$$\psi_\mu\begin{pmatrix} a & b \\ 0 & d \end{pmatrix} = \mu(ad).$$

We obtain the induced character

$$\psi_\mu^G = \mathrm{ind}_B^G(\psi_\mu).$$

Then ψ_μ^G is not simple. It contains $\mu \circ \det$, as one sees by Frobenius reciprocity:

$$\langle \mathrm{ind}_B^G \psi_\mu, \mu \circ \det \rangle_G = \langle \psi_\mu, \mu \circ \det \rangle_B = \frac{1}{\#(B)} \sum_{\beta \in B} |\mu \circ \det(\beta)|^2 = 1.$$

Characters $\chi = \psi_\mu^G - \mu \circ \det$ will be called of **second type**.

The values on the representatives of conjugacy classes are as follows.

Table 12.5(II)

χ	$\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$	$\begin{pmatrix} a & 1 \\ 0 & a \end{pmatrix}$	$\begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix} d \neq a$	$\alpha \in C - F^*$
$\psi_\mu^G - \mu \circ \det$	$q\mu(a)^2$	0	$\mu(ad)$	$-\mu \circ \det(\alpha)$

Actually, one computes the values of ψ_μ^G, and one then subtracts the value of $\theta \circ \det$. For this case and the next two cases, we use the formula for the induced function:

$$\mathrm{ind}_H^G(\varphi)(\alpha) = \frac{1}{\#(H)} \sum_{\beta \in G} \varphi_H(\beta\alpha\beta^{-1})$$

where φ_H is the function equal to φ on H and 0 outside H. An element of the center commutes with all $\beta \in G$, so for $\varphi = \psi_\mu$ the value of the induced character

on such an element is

$$\frac{\#(G)}{\#(B)}\mu(a)^2 = (q+1)\mu(a)^2,$$

which gives the stated value.

For an element $u = \begin{pmatrix} a & 1 \\ 0 & a \end{pmatrix}$, the only elements $\beta \in G$ such that $\beta u \beta^{-1}$ lies in B are the elements of B (by direct verification). It is then immediate that

$$\mathrm{ind}_B^G(\psi_\mu)\begin{pmatrix} a & 1 \\ 0 & a \end{pmatrix} = \mu(a)^2,$$

which yields the stated value for the character χ. Using Table 12.4, one finds at once that $\sum |\chi(\beta)|^2 = \#(G)$, and hence;

A character χ of second type is simple.

The table of values also shows that there are $q - 1$ characters of second type. The next two types deal especially with the Cartan subgroups.

Third type

$\psi : A \to \mathbf{C}^*$ denotes a homomorphism.

As mentioned following Proposition 12.3, the representative $w = w_A = w^{-1}$ for $N(A)/A$ is such that

$$w\begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}w = \begin{pmatrix} d & 0 \\ 0 & a \end{pmatrix} = \alpha^w \quad \text{if } \alpha = \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}.$$

Thus conjugation by w is an automorphism of order 2 on A. Let $[w]\psi$ be the conjugate character; that is, $([w]\psi)(\alpha) = \psi(w\alpha w) = \psi(\alpha^w)$ for $\alpha \in A$. Then $[w](\mu \circ \det) = \mu \circ \det$. The characters $\mu \circ \det$ on A are precisely those which are invariant under $[w]$. The others can be written in the form

$$\psi\begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix} = \psi_1(a)\psi_2(d),$$

with distinct characters $\psi_1, \psi_2 \colon F^* \to \mathbf{C}^*$. In light of the isomorphism $B/U \approx A$, we view ψ has a character on B. Then we form the induced character

$$\psi^G = \mathrm{ind}_B^G(\psi) = \mathrm{ind}_B^G([w]\psi).$$

With ψ such that $[w]\psi \neq \psi$, the characters $\chi = \psi^G$ will be said to be of the **third type**. Here is their table of values.

Table 12.5(III)

χ	$\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$	$\begin{pmatrix} a & 1 \\ 0 & a \end{pmatrix}$	$\alpha = \begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}_{d \neq a}$	$\alpha \in C - F^*$
ψ^G $\psi \neq [w]\psi$	$(q+1)\psi(a)$	$\psi(a)$	$\psi(\alpha) + \psi(\alpha^w)$	0

The first entry on central elements is immediate. For the second, we have already seen that if $\beta \in G$ is such that conjugating

$$\beta\begin{pmatrix} a & 1 \\ 0 & a \end{pmatrix}\beta^{-1} \in B,$$

then $\beta \in B$, and so the formula

$$\psi^G(\alpha) = \frac{1}{\#(B)} \sum_{\beta \in G} \psi_B(\beta\alpha\beta^{-1})$$

immediately gives the value of ψ^G on unipotent elements. For an element of A with $a \neq d$, there is the additional possibility of the normalizer of A with the elements w, and the value in the table then drops out from the formula. For elements of the non-split Cartan group, there is no element of G which conjugates them to elements of B, so the value in the last column is 0.

We claim that a character $\chi = \psi^G$ of third type is simple.

The proof again uses the test for simplicity, i.e. that $\sum |\chi(\beta)|^2 = \#(G)$. Observe that two elements α, $\alpha' \in A$ are in the same conjugacy class in G if and only if $\alpha' = \alpha$ or $\alpha' = [w]\alpha$. This is verified by brute force. Therefore, writing the sum $\sum |\psi^G(\beta)|^2$ for β in the various conjugacy classes, and using Table 12.4, we find:

$$\sum_{\beta \in G} |\psi^G(\beta)|^2 = (q+1)^2(q-1) + (q-1)(q^2-1) + (q^2+q) \sum_{\alpha \in (A-F^*)/w} |\psi(\alpha) + \psi(\alpha^w)|^2.$$

The third term can be written

$$\frac{1}{2}(q^2+q) \sum_{\alpha \in A\text{-}F^*} (\psi(\alpha) + \psi(\alpha^w))(\psi(\alpha^{-1}) + \psi(\alpha^{-w}))$$
$$= \frac{1}{2}(q^2+q) \sum_{\alpha \in A-F^*} (1 + 1 + \psi(\alpha^{1-w}) + \psi(\alpha^{w-1})).$$

We write the sum over $\alpha \in A - F^*$ as a sum for $\alpha \in A$ minus the sum for

$\alpha \in F^*$. If $\alpha \in F^*$ then $\alpha^{1-w} = \alpha^{w-1} = 1$. By assumption on ψ, the character

$$\alpha \mapsto \psi(\alpha^{1-w}) \text{ for } \alpha \in A$$

is non-trivial, and therefore the sum over $\alpha \in A$ is equal to 0. Therefore, putting these remarks together, we find that the third term is equal to

$$\frac{1}{2}(q^2 + q)[2(q-1)^2 - 2(q-1) - 2(q-1)] = q(q^2-1)(q-3).$$

Hence finally

$$\sum_{\beta \in G} |\psi^G(\beta)|^2 = (q+1)(q^2-1) + (q-1)(q^2-1) + q(q^2-1)(q-3)$$
$$= q(q-1)(q^2-1) = \#(G),$$

thus proving that ψ^G is simple.

Finally we observe that there are $\frac{1}{2}(q-1)(q-2)$ characters of third type. This is the number of characters ψ such that $[w]\psi \neq \psi$, divided by 2 because each pair ψ and $[w]\psi$ yields the same induced character ψ^G. The table of values shows that up to this coincidence, the induced characters are distinct.

Fourth type

$\theta : K^* \to \mathbf{C}^*$ denotes a homomorphism, which is viewed as a character on $C = C_K$.

By Proposition 12.3, there is an element $w \in N(C)$ but $w \notin C$, $w = w^{-1}$. Then

$$\alpha \mapsto w\alpha w = [w]\alpha$$

is an automorphism of C, but $x \mapsto wxw$ is also a field automorphism of $F[C] \approx K$ over F. Since $[K : F] = 2$, it follows that conjugation by w is the automorphism $\alpha \mapsto \alpha^q$. As a result we obtain the conjugate character $[w]\theta$ such that

$$([w]\theta)(\alpha) = \theta([w]\alpha) = \theta(\alpha^w),$$

and we get the induced character

$$\theta^G = \operatorname{ind}_C^G(\theta) = \operatorname{ind}_C^G([w]\theta).$$

Let $\mu : F^* \to \mathbf{C}^*$ denote a homomorphism as in the first type. Let:

$\lambda : F^+ \to \mathbf{C}^*$ be a *non-trivial* homomorphism.

$(\mu, \lambda) =$ the character on ZU such that

$$(\mu, \lambda)\left(\begin{pmatrix} a & ax \\ 0 & a \end{pmatrix}\right) = \mu(a)\lambda(x).$$

$(\mu, \lambda)^G = \operatorname{ind}_{ZU}^G(\mu, \lambda).$

A routine computation of the same nature that we have had previously gives the following values for the induced characters θ^G and $(\mu, \lambda)^G$.

χ	$\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$	$\begin{pmatrix} a & 1 \\ 0 & a \end{pmatrix}$	$\begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix} d \neq a$	$\alpha \in C - F^*$
θ^G	$(q^2 - q)\theta(a)$	0	0	$\theta(\alpha) + \theta(\alpha^w)$
$(\mu, \lambda)^G$	$(q^2 - 1)\mu(a)$	$-\mu(a)$	0	0

These are intermediate steps. Note that a direct computation using Frobenius reciprocity shows that θ^G occurs in the character $(\text{res}\,\theta, \lambda)^G$, where the restriction $\text{res}\,\theta$ is to the group F^*, so $\text{res}\,\theta$ is one of our characters μ. Thus we define:

$$\theta' = (\text{res}\,\theta, \lambda)^G - \theta^G = ([w]\theta)',$$

which is an effective character. A character θ' is said to be of **fourth type** if θ is such that $\theta \neq [w]\theta$. These are the characters we are looking for. Using the intermediate table of values, one then finds the table of values for those characters of fourth type.

Table 12.5(IV)

χ	$\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$	$\begin{pmatrix} a & 1 \\ 0 & a \end{pmatrix}$	$\begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix} d \neq a$	$\alpha \in C - F^*$
θ' $\theta \neq [w]\theta$	$(q - 1)\theta(a)$	$-\theta(a)$	0	$-\theta(\alpha) - \theta(\alpha^w)$

We claim that the characters θ' of fourth type are simple.

To prove this, we evaluate

$$\sum_{\beta \in G} |\theta'(\beta)|^2 = (q-1)^2(q-1) + (q-1)(q^2-1) + \frac{1}{2}(q^2 - q) \sum_{\alpha \in K^* - F^*} |\theta(\alpha) + \theta(\alpha^w)|^2.$$

We use the same type of expansion as for characters of third type, and the final value does turn out to be $\#(G)$, thus proving that θ' is simple.

The table also shows that there are $\frac{1}{2}\#(C - F^*) = \frac{1}{2}(q^2 - q)$ distinct characters of fourth type. We thus come to the end result of our computations.

Theorem 12.6. *The irreducible characters of $G = GL_2(F)$ are as follows.*

type		number of that type	dimension
I	$\mu \circ \det$	$q - 1$	1
II	$\psi_\mu^G - \mu \circ \det$	$q - 1$	q
III	ψ^G from pairs $\psi \neq [w]\psi$	$\frac{1}{2}(q-1)(q-2)$	$q + 1$
IV	θ' from pairs $\theta \neq [w]\theta$	$\frac{1}{2}(q-1)q$	$q - 1$

Proof. We have exhibited characters of four types. In each case it is immediate from our construction that we get the stated number of distinct characters of the given type. The dimensions as stated are immediately computed from the dimensions of induced characters as the index of the subgroup from which we induce, and on two occasions we have to subtract something which was needed to make the character of given type simple. The end result is the one given in the above table. The total number of listed characters is precisely equal to the number of classes in Table 12.4, and therefore we have found all the simple characters, thus proving the theorem.

EXERCISES

1. **The group S_3.** Let S_3 be the symmetric group on 3 elements,
   (a) Show that there are three conjugacy classes.
   (b) There are two characters of dimension 1, on S_3/A_3.
   (c) Let d_i $(i = 1, 2, 3)$ be the dimensions of the irreducible characters. Since $\sum d_i^2 = 6$, the third irreducible character has dimension 2. Show that the third representation can be realized by considering a cubic equation $X^3 + aX + b = 0$, whose Galois group is S_3 over a field k. Let V be the k-vector space generated by the roots. Show that this space is 2-dimensional and gives the desired representation, which remains irreducible after tensoring with k^a.
   (d) Let $G = S_3$. Write down an idempotent for each one of the simple components of $\mathbf{C}[G]$. What is the multiplicity of each irreducible representation of G in the regular representation on $\mathbf{C}[G]$?

2. **The groups S_4 and A_4.** Let S_4 be the symmetric group on 4 elements.
   (a) Show that there are 5 conjugacy classes.
   (b) Show that A_4 has a unique subgroup of order 4, which is not cyclic, and which is normal in S_4. Show that the factor group is isomorphic to S_3, so the representations of Exercise 1 give rise to representations of S_4.
   (c) Using the relation $\sum d_i^2 = \#(S_4) = 24$, conclude that there are only two other irreducible characters of S_4, each of dimension 3.
   (d) Let $X^4 + a_2X^2 + a_1X + a_0$ be an irreducible polynomial over a field k, with Galois group S_4. Show that the roots generate a 3-dimensional vector space V over k, and that the representation of S_4 on this space is irreducible, so we obtain one of the two missing representations.
   (e) Let ρ be the representation of (d). Define ρ' by

$$\rho'(\sigma) = \rho(\sigma) \text{ if } \sigma \text{ is even;}$$
$$\rho'(\sigma) = -\rho(\sigma) \text{ if } \sigma \text{ is odd.}$$

   Show that ρ' is also irreducible, remains irreducible after tensoring with k^{a}, and is non-isomorphic to ρ. This concludes the description of all irreducible representations of S_4.
   (f) Show that the 3-dimensional irreducible representations of S_4 provide an irreducible representation of A_4.
   (g) Show that all irreducible representations of A_4 are given by the representations in (f) and three others which are one-dimensional.

3. **The quaternion group**. Let $Q = \{\pm 1, \pm x, \pm y, \pm z\}$ be the quaternion group, with $x^2 = y^2 = z^2 = -1$ and $xy = -yx$, $xz = -zx$, $yz = -zy$.
   (a) Show that Q has 5 conjugacy classes.
   Let $A = \{\pm 1\}$. Then Q/A is of type (2, 2), and hence has 4 simple characters, which can be viewed as simple characters of Q.
   (b) Show that there is only one more simple character of Q, of dimension 2. Show that the corresponding representation can be given by a matrix representation such that

$$\rho(x) = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad \rho(y) = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \rho(z) = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix}.$$

   (c) Let $\mathbf{H}$ be the quaternion field, i.e. the algebra over $\mathbf{R}$ having dimension 4, with basis $\{1, x, y, z\}$ as in Exercise 3, and the corresponding relations as above. Show that $\mathbf{C} \otimes_{\mathbf{R}} \mathbf{H} \approx \mathrm{Mat}_2(\mathbf{C})$ (2×2 complex matrices). Relate this to (b).

4. Let S be a normal subgroup of G. Let ψ be a simple character of S over $\mathbf{C}$. Show that $\mathrm{ind}_S^G(\psi)$ is simple if and only if $\psi = [\sigma]\psi$ for all $\sigma \in S$.

5. Let G be a finite group and S a normal subgroup. Let ρ be an irreducible representation of G over $\mathbf{C}$. Prove that either the restriction of ρ to S has all its irreducible components S-isomorphic to each other, or there exists a proper subgroup H of G containing S and an irreducible representation θ of H such that $\rho \approx \mathrm{ind}_H^G(\theta)$.

6. **Dihedral group D_{2n}.** There is a group of order $2n$ (n even integer $\geqq 2$) generated by two elements σ, τ such that

$$\sigma^n = 1, \ \tau^2 = 1, \quad \text{and} \quad \tau\sigma\tau = \sigma^{-1}.$$

It is called the **dihedral group**.

(a) Show that there are four representations of dimension 1, obtained by the four possible values ± 1 for σ and τ.

(b) Let C_n be the cyclic subgroup of D_{2n} generated by σ. For each integer $r = 0, \ldots, n - 1$ let ψ_r be the character of C_n such that

$$\psi_r(\sigma) = \zeta^r \quad (\zeta = \text{prim. } n\text{-th root of unity})$$

Let χ_r be the induced character. Show that $\chi_r = \chi_{n-r}$.

(c) Show that for $0 < r < n/2$ the induced character χ_r is simple, of dimension 2, and that one gets thereby $\left(\frac{n}{2} - 1\right)$ distinct characters of dimension 2.

(d) Prove that the simple characters of (a) and (c) give all simple characters of D_{2n}.

7. Let G be a finite group, semidirect product of A, H where A is commutative and normal. Let $A^\wedge = \mathrm{Hom}(A, \mathbf{C}^*)$ be the dual group. Let G operate by conjugation on characters, so that for $\sigma \in G$, $a \in A$, we have

$$[\sigma]\psi(a) = \psi(\sigma^{-1}a\sigma).$$

Let $\psi_1, \ldots, \psi_r$ be representatives of the orbits of H in $A^\wedge$, and let $H_i (i = 1, \ldots, r)$ be the isotropy group of ψ_i. Let $G_i = AH_i$.

(a) For $a \in A$ and $h \in H_i$, define $\psi_i(ah) = \psi_i(a)$. Show that ψ_i is thus extended to a character on G_i.

Let θ be a simple representation of H_i (on a vector space over $\mathbf{C}$). From $H_i = G_i/A$, view θ as a simple representation of G_i. Let

$$\rho_{i,\theta} = \mathrm{ind}_{G_i}^G(\psi_i \otimes \theta).$$

(b) Show that $\rho_{i,\theta}$ is simple.

(c) Show that $\rho_{i,\theta} \approx \rho_{i',\theta'}$ implies $i = i'$ and $\theta \approx \theta'$.

(d) Show that every irreducible representation of G is isomorphic to some $\rho_{i,\theta}$

8. Let G be a finite group operating on a finite set S. Let $\mathbf{C}[S]$ be the vector space generated by S over $\mathbf{C}$. Let ψ be the character of the corresponding representation of G on $\mathbf{C}[S]$.

(a) Let $\sigma \in G$. Show that $\psi(\sigma)$ = number of fixed points of σ in S.

(b) Show that $\langle \psi, 1_G \rangle_G$ is the number of G-orbits in S.

9. Let A be a commutative subgroup of a finite group G. Show that every irreducible representation of G over $\mathbf{C}$ has dimension $\leqq (G : A)$.

10. Let $\mathbf{F}$ be a finite field and let $G = SL_2(\mathbf{F})$. Let B be the subgroup of G consisting of all matrices

$$\alpha = \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \in SL_2(\mathbf{F}), \text{ so } d = a^{-1}.$$

Let $\mu : \mathbf{F}^* \to \mathbf{C}^*$ be a homomorphism and let $\psi_\mu : B \to \mathbf{C}^*$ be the homomorphism such that $\psi_\mu(\alpha) = \mu(a)$. Show that the induced character $\mathrm{ind}_B^G(\psi_\mu)$ is simple if $\mu^2 \neq 1$.

11. Determine all simple characters of $SL_2(\mathbf{F})$, giving a table for the number of such characters, representatives for the conjugacy classes, as was done in the text for GL_2, over the complex numbers.

12. Observe that $A_5 \approx SL_2(\mathbf{F}_4) \approx PSL_2(\mathbf{F}_5)$. As a result, verify that there are 5 conjugacy classes, whose elements have orders 1, 2, 3, 5, 5 respectively, and write down explicitly the character table for A_5 as was done in the text for GL_2.

13. Let G be a p-group and let $G \to \mathrm{Aut}(V)$ be a representation on a finite dimensional vector space over a field of characteristic p. Assume that the representation is irreducible. Show that the representation is trivial, i.e. G acts as the identity on V.

14. Let G be a finite group and let C be a conjugacy class. Prove that the following two conditions are equivalent. They define what it means for the class to be **rational**.

    RAT 1. For all characters χ of G, $\chi(\sigma) \in \mathbf{Q}$ for $\sigma \in C$.

    RAT 2. For all $\sigma \in C$, and j prime to the order of σ, we have $\sigma^j \in C$.

15. Let G be a group and let H_1, H_2 be subgroups of finite index. Let ρ_1, ρ_2 be representations of H_1, H_2 on R-modules F_1, F_2 respectively. Let $M_G(F_1, F_2)$ be the R-module of functions $f : G \to \mathrm{Hom}_R(F_1, F_2)$ such that

$$f(h_1\sigma h_2) = \rho_2(h_2)f(\sigma)\rho_1(h_1)$$

for all $\sigma \in G$, $h_i \in H_i$ $(i = 1, 2)$. Establish an R-module isomorphism

$$\mathrm{Hom}_R(F_1^G, F_2^G) \xrightarrow{\approx} M_G(F_1, F_2).$$

By F_i^G we have abbreviated $\mathrm{ind}_{H_i}^G(F_i)$.

16. (a) Let G_1, G_2 be two finite groups with representations on $\mathbf{C}$-spaces E_1, E_2. Let $E_1 \otimes E_2$ be the usual tensor product over $\mathbf{C}$, but now prove that there is an action of $G_1 \times G_2$ on this tensor product such that

$$(\sigma_1, \sigma_2)(x \otimes y) = \sigma_1 x \otimes \sigma_2 y \text{ for } \sigma_1 \in G_1, \sigma_2 \in G_2.$$

    This action is called the **tensor product** of the other two. If ρ_1, ρ_2 are the representations of G_1, G_2 on E_1, E_2 respectively, then their tensor product is denoted by $\rho_1 \otimes \rho_2$. Prove: If ρ_1, ρ_2 are irreducible then $\rho_2 \otimes \rho_2$ is also irreducible. [*Hint*: Use Theorem 5.17.]

    (b) Let χ_1, χ_2 be the characters of ρ_1, ρ_2 respectively. Show that $\chi_1 \otimes \chi_2$ is the character of the tensor product. By definition,

$$\chi_1 \otimes \chi_2(\sigma_1, \sigma_2) = \chi_1(\sigma_1)\,\chi_2(\sigma_2).$$

17. With the same notation as in Exercise 16, show that every irreducible representation of $G_1 \times G_2$ over $\mathbf{C}$ is isomorphic to a tensor product representation as in Exercise 16. [*Hint*: Prove that if a character is orthogonal to all the products $\chi_1 \otimes \chi_2$ of Exercise 16(b) then the character is 0.]

Tensor product representations

18. Let P be the non-commutative polynomial algebra over a field k, in n variables. Let $x_1, \ldots, x_r$ be distinct elements of P_1 (i.e. linear expressions in the variables $t_1, \ldots, t_n$)

and let $a_1, \ldots, a_r \in k$. If

$$a_1 x_1^\nu + \cdots + a_r x_r^\nu = 0$$

for all integers $\nu = 1, \ldots, r$ show that $a_i = 0$ for $i = 1, \ldots, r$. [*Hint*: Take the homomorphism on the commutative polynomial algebra and argue there.]

19. Let G be a finite set of endomorphisms of a finite-dimensional vector space E over the field k. For each $\sigma \in G$, let c_σ be an element of k. Show that if

$$\sum_{\sigma \in G} c_\sigma T^r(\sigma) = 0$$

for all integers $r \geqq 1$, then $c_\sigma = 0$ for all $\sigma \in G$. [*Hint*: Use the preceding exercise, and Proposition 7.2 of Chapter XVI.]

20. **(Steinberg)**. Let G be a finite monoid, and $k[G]$ the monoid algebra over a field k. Let $G \to \text{End}_k(E)$ be a faithful representation (i.e. injective), so that we identify G with a multiplicative subset of $\text{End}_k(E)$. Show that T^r induces a representation of G on $T^r(E)$, whence a representation of $k[G]$ on $T^r(E)$ by linearity. If $\alpha \in k[G]$ and if $T^r(\alpha) = 0$ for all integers $r \geqq 1$, show that $\alpha = 0$. [*Hint*: Apply the preceding exercise.]

21. **(Burnside)**. Deduce from Exercise 20 the following theorem of Burnside: Let G be a finite group, k a field of characteristic prime to the order of G, and E a finite dimensional (G, k)-space such that the representation of G is faithful. Then every irreducible representation of G appears with multiplicity $\geqq 1$ in some tensor power $T^r(E)$.

22. Let $X(G)$ be the character ring of a finite group G, generated over $\mathbf{Z}$ by the simple characters over $\mathbf{C}$. Show that an element $f \in X(G)$ is an effective irreducible character if and only if $\langle f, f \rangle_G = 1$ and $f(1) \geqq 0$.

23. In this exercise, we assume the next chapter on alternating products. Let ρ be an irreducible representation of G on a vector space E over $\mathbf{C}$. Then by functoriality we have the corresponding representations $S^r(\rho)$ and $\bigwedge^r(\rho)$ on the r-th symmetric power and r-th alternating power of E over $\mathbf{C}$. If χ is the character of ρ, we let $S^r(\chi)$ and $\bigwedge^r(\chi)$ be the characters of $S^r(\rho)$ and $\bigwedge^r(\rho)$ respectively, on $S^r(E)$ and $\bigwedge^r(E)$. Let t be a variable and let

$$\sigma_t(\chi) = \sum_{r=0}^{\infty} S^r(\chi) t^r, \quad \lambda_t(\chi) = \sum_{r=0}^{\infty} \textstyle\bigwedge^r(\chi) t^r.$$

(a) Comparing with Exercise 24 of Chapter XIV, prove that for $x \in G$ we have

$$\sigma_t(\chi)(x) = \det(I - \rho(x)t)^{-1} \quad \text{and} \quad \lambda_t(\chi)(x) = \det(I + \rho(x)t).$$

(b) For a function f on G define $\Psi^n(f)$ by $\Psi^n(f)(x) = f(x^n)$. Show that

$$-\frac{d}{dt}\log \sigma_t(\chi) = \sum_{n=1}^{\infty} \Psi^n(\chi) t^n \quad \text{and} \quad -\frac{d}{dt}\log \lambda_{-t}(\chi) = \sum_{n=1}^{\infty} \Psi^n(\chi) t^n.$$

(c) Show that

$$nS^n(\chi) = \sum_{r=1}^{n} \Psi^r(\chi) S^{n-r}(\chi) \quad \text{and} \quad n\textstyle\bigwedge^n(\chi) = \sum_{r=1}^{\infty} (-1)^{r-1} \Psi^r(\chi) \textstyle\bigwedge^{n-r}(\chi).$$

24. Let χ be a simple character of G. Prove that $\Psi^n(\chi)$ is also simple. (The characters are over $\mathbf{C}$.)

25. We now assume that you know §3 of Chapter XX.
    (a) Prove that the Grothendieck ring defined there for $\text{Mod}_{\mathbf{C}}(G)$ is naturally isomorphic to the character ring $X(G)$.
    (b) Relate the above formulas with Theorem 3.12 of Chapter XX.
    (c) Read Fulton-Lang's *Riemann-Roch Algebra*, Chapter I, especially §6, and show that $X(G)$ is a λ-ring, with Ψ^n as the Adams operations.

Note. For further connections with homology and the cohomology of groups, see Chapter XX, §3, and the references given at the end of Chapter XX, §3.

26. The following formalism is the analogue of Artin's formalism of L-series in number theory. Cf. Artin's "Zur Theorie der L-Reihen mit allgemeinen Gruppencharakteren", Collected papers, and also S. Lang, "L-series of a covering", *Proc. Nat. Acad. Sc. USA* (1956). For the Artin formalism in a context of analysis, see J. Jorgenson and S. Lang, "Artin formalism and heat kernels", *J. reine angew. Math.* **447** (1994) pp. 165–200.

We consider a category with objects $\{U\}$. As usual, we say that a finite group G operates on U if we are given a homomorphism $\rho : G \to \text{Aut}(U)$. We then say that U is a G-object, and also that ρ is a representation of G in U. We say that G operates trivially if $\rho(G) = \text{id}$. For simplicity, we omit the ρ from the notation. By a G-morphism $f: U \to V$ between G-objects, one means a morphism such that $f \circ \sigma = \sigma \circ f$ for all $\sigma \in G$.

We shall assume that for each G-object U there exists an object U/G on which G operates trivially, and a G-morphism $\pi_{U,G} : U \to U/G$ having the following universal property: If $f : U \to U'$ is a G-morphism, then there exists a unique morphism

$$f/G : U/G \to U'/G$$

making the following diagram commutative:

$$\begin{array}{ccc} U & \xrightarrow{f} & U' \\ \downarrow & & \downarrow \\ U/G & \xrightarrow[f/G]{} & U'/G \end{array}$$

In particular, if H is a normal subgroup of G, show that G/H operates in a natural way on U/H.

Let k be an algebraically closed field of characteristic 0. We assume given a functor E from our category to the category of finite dimensional k-spaces. If U is an object in our category, and $f : U \to U'$ is a morphism, then we get a homomorphism

$$E(f) = f_* : E(U) \to E(U').$$

(The reader may keep in mind the special case when we deal with the category of reasonable topological spaces, and E is the homology functor in a given dimension.)

If G operates on U, then we get an operation of G on $E(U)$ by functoriality.

Let U be a G-object, and $F : U \to U$ a G-morphism. If $P_F(t) = \prod (t - \alpha_i)$ is the characteristic polynomial of the linear map $F_* : E(U) \to E(U)$, we define

$$Z_F(t) = \prod (1 - \alpha_i t),$$

and call this the zeta function of F. If F is the identity, then $Z_F(t) = (1 - t)^{B(U)}$ where we define $B(U)$ to be $\dim_k E(U)$.

Let χ be a simple character of G. Let d_χ be the dimension of the simple representation of G belonging to χ, and $n = \operatorname{ord}(G)$. We define a linear map on $E(U)$ by letting

$$e_\chi = \frac{d_\chi}{n} \sum_{\sigma \in G} \chi(\sigma^{-1})\sigma_*.$$

Show that $e_\chi^2 = e_\chi$, and that for any positive integer μ we have $(e_\chi \circ F_*)^\mu = e_\chi \circ F_*^\mu$. If $P_\chi(t) = \prod (t - \beta_j(\chi))$ is the characteristic polynomial of $e_\chi \circ F_*$, define

$$L_F(t, \chi, U/G) = \prod (1 - \beta_j(\chi)t).$$

Show that the logarithmic derivative of this function is equal to

$$-\frac{1}{N} \sum_{\mu=1}^{\infty} \operatorname{tr}(e_\chi \circ F_*^\mu)t^{\mu-1}.$$

Define $L_F(t, \chi, U/G)$ for any character χ by linearity. If we write $V = U/G$ by abuse of notation, then we also write $L_F(t, \chi, U/V)$. Then for any χ, χ' we have by definition,

$$L_F(t, \chi + \chi', U/V) = L_F(t, \chi, U/V)L_F(t, \chi', U/V).$$

We make one additional assumption on the situation:
Assume that the characteristic polynomial of

$$\frac{1}{n} \sum_{\sigma \in G} \sigma_* \circ F_*$$

is equal to the characteristic polynomial of F/G on $E(U/G)$. Prove the following statement:

(a) If $G = \{1\}$ then

$$L_F(t, 1, U/U) = Z_F(t).$$

(b) Let $V = U/G$. Then

$$L_F(t, 1, U/V) = Z_F(t).$$

(c) Let H be a subgroup of G and let ψ be a character of H. Let $W = U/H$, and let ψ^G be the induced character from H to G. Then

$$L_F(t, \psi, U/W) = L_F(t, \psi^G, U/V).$$

(d) Let H be normal in G. Then G/H operates on $U/H = W$. Let ψ be a character of G/H, and let χ be the character of G obtained by composing ψ with the canonical map $G \to G/H$. Let $\varphi = F/H$ be the morphism induced on

$$U/H = W.$$

Then

$$L_\varphi(t, \psi, W/V) = L_F(t, \chi, U/V).$$

(e) If $V = U/G$ and $B(V) = \dim_k E(V)$, show that $(1 - t)^{B(V)}$ divides $(1 - t)^{B(U)}$. Use the regular character to determine a factorization of $(1 - t)^{B(U)}$.

27. Do this exercise after you have read some of Chapter VII. The point is that for fields of characteristic not dividing the order of the group, the representations can be obtained by "reducing modulo a prime". Let G be a finite group and let p be a prime not dividing the order of G. Let F be a finite extension of the rationals with ring of algebraic integers $\mathfrak{o}_F$. Suppose that F is sufficiently large so that all F-irreducible representations of G remain irreducible when tensored with $\mathbf{Q}^{\mathrm{a}} = F^{\mathrm{a}}$. Let $\mathfrak{p}$ be a prime of $\mathfrak{o}_F$ lying above p, and let $\mathfrak{o}_\mathfrak{p}$ be the corresponding local ring.
    (a) Show that an irreducible (G, F)-space V can be obtained from a $(G, \mathfrak{o}_\mathfrak{p})$-module E free over $\mathfrak{o}_\mathfrak{p}$, by extending the base from $\mathfrak{o}_\mathfrak{p}$ to F, i.e. by tensoring so that $V = E \otimes F$ (tensor product over $\mathfrak{o}_\mathfrak{p}$).
    (b) Show that the reduction mod $\mathfrak{p}$ of E is an irreducible representation of G in characteristic p. In other words, let $k = \mathfrak{o}/\mathfrak{p} = \mathfrak{o}_\mathfrak{p}/\mathfrak{m}_\mathfrak{p}$ where $\mathfrak{m}_\mathfrak{p}$ is the maximal ideal of $\mathfrak{o}_\mathfrak{p}$. Let $E(\mathfrak{p}) = E \otimes k$ (tensor product over $\mathfrak{o}_\mathfrak{p}$). Show that G operates on $E(\mathfrak{p})$ in a natural way, and that this representation is irreducible. In fact, if χ is the character of G on V, show that χ is also the character on E, and that χ mod $\mathfrak{m}_\mathfrak{p}$ is the character on $E(\mathfrak{p})$.
    (c) Show that all irreducible characters of G in characteristic p are obtained as in (b).

CHAPTER XIX

The Alternating Product

The alternating product has applications throughout mathematics. In differential geometry, one takes the maximal alternating product of the tangent space to get a canonical line bundle over a manifold. Intermediate alternating products give rise to differential forms (sections of these products over the manifold). In this chapter, we give the algebraic background for these constructions.

For a reasonably self-contained treatment of the action of various groups of automorphisms of bilinear forms on tensor and alternating algebras, together with numerous classical examples, I refer to:

R. HOWE, Remarks on classical invariant theory, *Trans. AMS* **313** (1989), pp. 539–569

§1 DEFINITION AND BASIC PROPERTIES

Consider the category of modules over a commutative ring R.

We recall that an r-multilinear map $f: E^{(r)} \to F$ is said to be **alternating** if $f(x_1, \ldots, x_r) = 0$ whenever $x_i = x_j$ for some $i \neq j$.

Let $\mathfrak{a}_r$ be the submodule of the tensor product $T^r(E)$ generated by all elements of type

$$x_1 \otimes \cdots \otimes x_r$$

where $x_i = x_j$ for some $i \neq j$. We define

$$\bigwedge^r(E) = T^r(E)/\mathfrak{a}_r.$$

Then we have an r-multilinear map $E^{(r)} \to \bigwedge^r(E)$ (called canonical) obtained

from the composition

$$E^{(r)} \to T^r(E) \to T^r(E)/\mathfrak{a}_r = \bigwedge^r(E).$$

It is clear that our map is alternating. *Furthermore, it is universal with respect to r-multilinear alternating maps on E.* In other words, if $f : E^{(r)} \to F$ is such a map, there exists a unique linear map $f_* : \bigwedge^r(E) \to F$ such that the following diagram is commutative:

$$\begin{array}{ccc} & & \bigwedge^r(E) \\ & \nearrow & \downarrow f_* \\ E^{(r)} & & \\ & \searrow_f & F \end{array}$$

Our map f_* exists because we can first get an induced map $T^r(E) \to F$ making the following diagram commutative:

$$\begin{array}{ccc} & & T^r(E) \\ & \nearrow & \downarrow \\ E^{(r)} & & \\ & \searrow_f & F \end{array}$$

and this induced map vanishes on $\mathfrak{a}_r$, hence inducing our f_*.

The image of an element $(x_1, \ldots, x_r) \in E^{(r)}$ in the canonical map into $\bigwedge^r(E)$ will be denoted by $x_1 \wedge \cdots \wedge x_r$. It is also the image of $x_1 \otimes \cdots \otimes x_r$ in the factor homomorphism $T^r(E) \to \bigwedge^r(E)$.

In this way, $\bigwedge^r$ becomes a functor, from modules to modules. Indeed, let $u: E \to F$ be a homomorphism. Given elements $x_1, \ldots, x_r \in E$, we can map

$$(x_1, \ldots, x_r) \mapsto u(x_1) \wedge \cdots \wedge u(x_r) \in \bigwedge^r(F).$$

This map is multilinear alternating, and therefore induces a homomorphism

$$\bigwedge^r(u): \bigwedge^r(E) \to \bigwedge^r(F).$$

The association $u \mapsto \bigwedge^r(u)$ is obviously functorial.

Example. Open any book on differential geometry (complex or real) and you will see an application of this construction when E is the tangent space of a point on a manifold, or the dual of the tangent space. When taking the dual, the construction gives rise to differential forms.

We let $\bigwedge(E)$ be the direct sum

$$\bigwedge(E) = \bigoplus_{r=0}^{\infty} \bigwedge^r(E).$$

We shall make $\bigwedge(E)$ into a graded R-algebra and call it the **alternating algebra** of E, or also the **exterior algebra**, or the **Grassmann algebra**. We shall first discuss the general situation, with arbitrary graded rings.

Let G be an additive monoid again, and let $A = \bigoplus_{r \in G} A_r$ be a G-graded R-algebra. Suppose given for each A_r a submodule $\mathfrak{a}_r$, and let $\mathfrak{a} = \bigoplus_{r \in G} \mathfrak{a}_r$. Assume that $\mathfrak{a}$ is an ideal of A. Then $\mathfrak{a}$ is called a **homogeneous ideal**, and we can define a graded structure on $A/\mathfrak{a}$. Indeed, the bilinear map

$$A_r \times A_s \to A_{r+s}$$

sends $\mathfrak{a}_r \times A_s$ into $\mathfrak{a}_{r+s}$ and similarly, sends $A_r \times \mathfrak{a}_s$ into $\mathfrak{a}_{r+s}$. Thus using representatives in A_r, A_s respectively, we can define a bilinear map

$$A_r/\mathfrak{a}_r \times A_s/\mathfrak{a}_s \to A_{r+s}/\mathfrak{a}_{r+s},$$

and thus a bilinear map $A/\mathfrak{a} \times A/\mathfrak{a} \to A/\mathfrak{a}$, which obviously makes $A/\mathfrak{a}$ into a graded R-algebra.

We apply this to $T^r(E)$ and the modules $\mathfrak{a}_r$ defined previously. If

$$x_i = x_j \quad (i \neq j)$$

in a product $x_1 \wedge \cdots \wedge x_r$, then for any $y_1, \ldots, y_s \in E$ we see that

$$x_1 \wedge \cdots \wedge x_r \wedge y_1 \wedge \cdots \wedge y_s$$

lies in $\mathfrak{a}_{r+s}$, and similarly for the product on the left. Hence the direct sum $\bigoplus \mathfrak{a}_r$ is an ideal of $T(E)$, and we can define an R-algebra structure on $T(E)/\mathfrak{a}$. The product on homogeneous elements is given by the formula

$$((x_1 \wedge \cdots \wedge x_r), (y_1 \wedge \cdots \wedge y_s)) \mapsto x_1 \wedge \cdots \wedge x_r \wedge y_1 \wedge \cdots \wedge y_s.$$

We use the symbol $\wedge$ also to denote the product in $\bigwedge(E)$. This product is called the **alternating product** or **exterior product**. If $x \in E$ and $y \in E$, then $x \wedge y = -y \wedge x$, as follows from the fact that $(x + y) \wedge (x + y) = 0$.

We observe that $\bigwedge$ is a functor from the category of modules to the category of graded R-algebras. To each linear map $f : E \to F$ we obtain a map

$$\bigwedge(f) : \bigwedge(E) \to \bigwedge(F)$$

which is such that for $x_1, \ldots, x_r \in E$ we have

$$\bigwedge(f)(x_1 \wedge \cdots \wedge x_r) = f(x_1) \wedge \cdots \wedge f(x_r).$$

Furthermore, $\bigwedge(f)$ is a homomorphism of graded R-algebras.

Proposition 1.1. *Let E be free of dimension n over R. If $r > n$ then $\bigwedge^r(E) = 0$. Let $\{v_1, \ldots, v_n\}$ be a basis of E over R. If $1 \leqq r \leqq n$, then $\bigwedge^r(E)$ is free over R, and the elements*

$$v_{i_1} \wedge \cdots \wedge v_{i_r}, \qquad i_1 < \cdots < i_r$$

form a basis of $\bigwedge^r(E)$ over k. We have

$$\dim_R \bigwedge^r(E) = \binom{n}{r}.$$

Proof. We shall first prove our assertion when $r = n$. Every element of E can be written in the form $\sum a_i v_i$, and hence using the formula $x \wedge y = -y \wedge x$ we conclude that $v_1 \wedge \cdots \wedge v_n$ generates $\bigwedge^n(E)$. On the other hand, we know from the theory of determinants that given $a \in R$, there exists a unique multilinear alternating form f_a on E such that

$$f_a(v_1, \ldots, v_n) = a.$$

Consequently, there exists a unique linear map

$$\bigwedge^n(E) \to R$$

taking the value a on $v_1 \wedge \cdots \wedge v_n$. From this it follows at once that $v_1 \wedge \cdots \wedge v_n$ is a basis of $\bigwedge^n(E)$ over R.

We now prove our statement for $1 \leqq r \leqq n$. Suppose that we have a relation

$$0 = \sum a_{(i)} v_{i_1} \wedge \cdots \wedge v_{i_r}$$

with $i_1 < \cdots < i_r$ and $a_{(i)} \in R$. Select any r-tuple $(j) = (j_1, \ldots, j_r)$ such that $j_1 < \cdots < j_r$ and let $j_{r+1}, \ldots, j_n$ be those values of i which do not appear among $(j_1, \ldots, j_r)$. Take the alternating product with $v_{j_{r+1}} \wedge \cdots \wedge v_{j_n}$. Then we shall have alternating products in the sum with repeated components in all the terms except the (j)-term, and thus we obtain

$$0 = a_{(j)} v_{j_1} \wedge \cdots \wedge v_{j_r} \wedge \cdots \wedge v_{j_n}.$$

Reshuffling $v_{j_1} \wedge \cdots \wedge v_{j_n}$ into $v_1 \wedge \cdots \wedge v_n$ simply changes the right-hand side by a sign. From what we proved at the beginning of this proof, it follows that $a_{(j)} = 0$. Hence we have proved our assertion for $1 \leqq r \leqq n$.

When $r = 0$, we deal with the empty product, and 1 is a basis for $\bigwedge^0(E) = R$ over R. We leave the case $r > n$ as a trivial exercise to the reader.

The assertion concerning the dimension is trivial, considering that there is a bijection between the set of basis elements, and the subsets of the set of integers $(1, \ldots, n)$.

Remark. It is possible to give the first part of the proof, for $\bigwedge^n(E)$, without assuming known the existence of determinants. One must then show that $\mathfrak{a}_n$ admits a 1-dimensional complementary submodule in $T^n(E)$. This can be done by simple means, which we leave as an exercise which the reader can look up in the more general situation of §4. When R is a field, this exercise is even more trivial, since one can verify at once that $v_1 \otimes \cdots \otimes v_n$ does not lie in $\mathfrak{a}_n$. This alternative approach to the theorem then proves the existence of determinants.

Proposition 1.2. *Let*

$$0 \to E' \to E \to E'' \to 0$$

be an exact sequence of free R-modules of finite ranks r, n, and s respectively. Then there is a natural isomorphism

$$\varphi : \bigwedge^r E' \otimes \bigwedge^s E'' \to \bigwedge^n E.$$

This isomorphism is the unique isomorphism having the following property. For elements $v_1, \ldots, v_r \in E'$ *and* $w_1, \ldots, w_s \in E''$, *let* $u_1, \ldots, u_s$ *be liftings of* $w_1, \ldots, w_s$ *in E. Then*

$$\varphi((v_1 \wedge \cdots \wedge v_r) \otimes (w_1 \wedge \cdots \wedge w_s)) = v_1 \wedge \cdots \wedge v_r \wedge u_1 \wedge \cdots \wedge u_s.$$

Proof. The proof proceeds in the usual two steps. First one shows the existence of a homomorphism φ having the desired effect. The value on the right of the above formula is independent of the choice of $u_1, \ldots, u_s$ lifting $w_1, \ldots, w_s$ by using the alternating property, so we obtain a homomorphism φ. Selecting in particular $\{v_1, \ldots, v_r\}$ and $\{w_1, \ldots, w_s\}$ to be bases of E' and E'' respectively, one then sees that φ is both injective and surjective. We leave the details to the reader.

Given a free module E of rank n, we define its **determinant** to be

$$\det E = \bigwedge^{\max} E = \bigwedge^n E.$$

Then Proposition 1.2 may be reformulated by the isomorphism formula

$$\det(E') \otimes \det(E'') \approx \det(E).$$

If $R = k$ is a field, then we may say that det is an Euler-Poincaré map on the category of finite dimensional vector spaces over k.

Example. Let V be a finite dimensional vector space over $\mathbf{R}$. By a **volume** on V we mean a norm $\| \ \|$ on $\det V$. Since V is finite dimensional, such a norm is equivalent to assigning a positive number c to a given basis of $\det(V)$. Such a basis can be expressed in the form $e_1 \wedge \cdots \wedge e_n$, where $\{e_1, \ldots, e_n\}$ is a basis of V. Then for $a \in \mathbf{R}$ we have

$$\|ae_1 \wedge \cdots \wedge e_n\| = |a|c.$$

In analysis, given a volume as above, one then defines a Haar measure μ on V by defining the measure of a set S to be

$$\mu(S) = \int_S \|e_1 \wedge \cdots \wedge e_n\| \, dx_1 \cdots dx_n,$$

where $x_1, \ldots, x_n$ are the coordinates on V with respect to the above basis. As an exercise, show that the expression on the right is the independent of the choice of basis.

Proposition 1.2 is a special case of the following more general situation. We consider again an exact sequence of free R-modules of finite rank as above. With respect to the submodule E' of E, we define

$$\bigwedge_i^n E = \text{submodule of } \bigwedge^n E \text{ generated by all elements}$$

$$x'_1 \wedge \cdots \wedge x'_i \wedge y_{i+1} \wedge \cdots \wedge y_n$$

with $x'_1, \ldots, x'_i \in E'$ viewed as submodule of E.

Then we have a filtration

$$\bigwedge_i^n E \supset \bigwedge_{i+1}^n E.$$

Proposition 1.3. *There is a natural isomorphism*

$$\bigwedge^i E' \otimes \bigwedge^{n-i} E'' \to \bigwedge_i^n E / \bigwedge_{i+1}^n E.$$

Proof. Let $x''_1, \ldots, x''_{n-i}$ be elements of E'', and lift them to elements $y_1, \ldots, y_{n-i}$ of E. We consider the map

$$(x'_1, \ldots, x'_i, x''_1, \ldots, x''_{n-i}) \mapsto x'_1 \wedge \cdots \wedge x'_i \wedge y_1 \wedge \cdots \wedge y_{n-i}$$

with the right-hand side taken mod $\bigwedge_{i+1}^n E$. Then it is immediate that this map factors through

$$\bigwedge^i E' \otimes \bigwedge^{n-i} E'' \to \bigwedge_i^n E / \bigwedge_{i+1}^n E,$$

and picking bases shows that one gets an isomorphism as desired.

In a similar vein, we have:

Proposition 1.4. *Let $E = E' \oplus E''$ be a direct sum of finite free modules. Then for every positive integer n, we have a module isomorphism*

$$\bigwedge^n E \approx \bigoplus_{p+q=n} \bigwedge^p E' \otimes \bigwedge^q E''.$$

In terms of the alternating algebras, we have an isomorphism

$$\bigwedge E \approx \bigwedge E' \otimes_{su} \bigwedge E''.$$

where $\otimes_{su}$ *is the superproduct of graded algebras.*

Proof. Each natural injection of E' and E'' into E induces a natural map on the alternating algebras, and so gives the homomorphism

$$\bigwedge E' \otimes \bigwedge E'' \to \bigwedge E,$$

which is graded, i.e. for $p = 0, \dots, n$ we have

$$\bigwedge^p E' \otimes \bigwedge^{n-p} E'' \to \bigwedge^n E.$$

To verify that this yields the desired isomorphism, one can argue by picking bases, which we leave to the reader. The anti-commutation rule of the alternating product immediately shows that the isomorphism is an algebra isomorphism for the super product $\bigwedge E' \otimes_{su} \bigwedge E''$.

We end this section with comments on duality. In Exercise 3, you will prove:

Proposition 1.5. *Let E be free of rank n over R. For each positive integer r, we have a natural isomorphism*

$$\bigwedge^r(E^\vee) \approx \bigwedge^r(E)^\vee.$$

The isomorphism is explicitly described in that exercise. A more precise property than "natural" would be that the isomorphism is functorial with respect to the category whose objects are finite free modules over R, and whose morphisms are isomorphisms.

Examples. Let L be a free module over R of rank 1. We have the dual module $L^\vee = \mathrm{Hom}_R(L, R)$, which is also free of the same rank. For a positive integer m, we define

$$L^{\otimes -m} = (L^\vee)^{\otimes m} = L^\vee \otimes \cdots \otimes L^\vee \text{ (tensor product taken } m \text{ times).}$$

Thus we have defined the tensor product of a line with itself for negative integers. We define $L^{\otimes 0} = R$. You can easily verify that the rule

$$L^{\otimes p} \otimes L^{\otimes q} \approx L^{\otimes (p+q)}$$

holds for all integers $p, q \in \mathbf{Z}$, with a natural isomorphism. In particular, if $q = -p$ then we get R itself on the right-hand side.

Now let $\mathbf{E}$ be an exact sequence of free modules:

$$\mathbf{E} : 0 \to E_0 \to E_1 \to \cdots \to E_m \to 0.$$

We define the **determinant** of this exact sequence to be

$$\det(\mathbf{E}) = \bigotimes \det(E_i)^{\otimes(-1)^i}.$$

As an exercise, prove that $\det(\mathbf{E})$ has a natural isomorphism with R, functorial with respect to isomorphisms of exact sequences.

Examples. Determinants of vector spaces or free modules occur in several branches of mathematics, e.g. complexes of partial differential operators, homology theories, the theory of determinant line bundles in algebraic geometry, etc. For instance, given a non-singular projective variety V over $\mathbf{C}$, one defines the **determinant of cohomology** of V to be

$$\det H(V) = \bigotimes \det H^i(V)^{\otimes(-1)^i},$$

where $H^i(V)$ are the cohomology groups. Then $\det H(V)$ is a one-dimensional vector space over $\mathbf{C}$, but there is no natural identification of this vector space with $\mathbf{C}$, because *a priori* there is no natural choice of a basis. For a notable application of the determinant of cohomology, following work of Faltings, see Deligne, Le determinant de la cohomologie, in Ribet, K. (ed.), *Current Trends in Arithmetical Algebraic Geometry*, Proc. Arcata 1985. (*Contemporary Math.* vol **67**, AMS (1985), pp. 93–178.)

§2. FITTING IDEALS

Certain ideals generated by determinants are coming more and more into use, in several branches of algebra and algebraic geometry. Therefore I include this section which summarizes some of their properties. For a more extensive account, see Northcott's book *Finite Free Resolutions* which I have used, as well as the appendix of the paper by Mazur-Wiles: "Class Fields of abelian extensions of $\mathbf{Q}$," which they wrote in a self-contained way. (*Invent. Math.* **76** (1984), pp. 179–330.)

Let R be a commutative ring. Let A be a $p \times q$ matrix and B a $q \times s$ matrix with coefficients in R. Let $r \geqq 0$ be an integer. We define the **determinant ideal** $I_r(A)$ to be the ideal generated by all determinants of $r \times r$ submatrices of A. This ideal may also be described as follows. Let S_r^p be the set of sequences

$$J = (j_1, \ldots, j_r) \text{ with } 1 \leqq j_1 < j_2 < \cdots < j_r \leqq p.$$

Let $A = (a_{ij})$. Let $1 \leqq r \leqq \min(p, q)$. Let $K = (k_1, \ldots, k_r)$ be another element of S_r^p. We *define*

$$A_{JK}^{(r)} = \begin{vmatrix} a_{j_1k_1} & a_{j_1k_2} & \cdots & a_{j_1k_r} \\ a_{j_2k_1} & a_{j_2k_2} & \cdots & a_{j_2k_r} \\ \vdots & \vdots & & \vdots \\ a_{j_rk_1} & a_{j_rk_2} & \cdots & a_{j_rk_r} \end{vmatrix}$$

where the vertical bars denote the determinant. With J, K ranging over S_r^p we may view $A_{JK}^{(r)}$ as the JK-component of a matrix $A^{(r)}$ which we call the r-th **exterior power** of A.

One may also describe the matrix as follows. Let $\{e_1, \ldots, e_p\}$ be a basis of R^p and $\{u_1, \ldots, u_q\}$ a basis of R^q. Then the elements

$$e_{j_1} \wedge \cdots \wedge e_{j_r} \qquad (j_1 < j_2 < \cdots < j_r)$$

form a basis for $\bigwedge^r R^p$ and similarly for a basis of $\bigwedge^r R^q$. We may view A as a linear map of R^p into R^q, and the matrix $A^{(r)}$ is then the matrix representing the exterior power $\bigwedge^r A$ viewed as a linear map of $\bigwedge^r R^p$ into $\bigwedge^r R^q$. On the whole, this interpretation will not be especially useful for certain computations, but it does give a slightly more conceptual context for the exterior power. Just at the beginning, this interpretation allows for an immediate proof of Proposition 2.1.

For $r = 0$ we define $A^{(0)}$ to be the 1×1 matrix whose single entry is the unit element of R. We also note that $A^{(1)} = A$.

Proposition 2.1. *Let A be a $p \times q$ matrix and B a $q \times s$ matrix. Then*

$$(AB)^{(r)} = A^{(r)}B^{(r)} \quad \textit{for} \quad r \geqq 0.$$

If one uses the alternating products as mentioned above, the proof simply says that the matrix of the composite of linear maps with respect to fixed bases is the product of the matrices. If one does not use the alternating products, then one can prove the proposition by a direct computation which will be left to the reader.

We have formed a matrix whose entries are indexed by a finite set S_r^p. For any finite set S and doubly indexed family (c_{JK}) with J, $K \in S$ we may also define the **determinant** as

$$\det(c_{JK}) = \sum_{\sigma} \epsilon(\sigma)\left(\prod_{J \in S} c_{J,\sigma(J)}\right)$$

where σ ranges over all permutations of the set.

For $r \geqq 0$ we define the **determinant ideal** $I_r(A)$ to be the ideal generated by all the components of $A^{(r)}$, or equivalently by all $r \times r$ subdeterminants of A. We have by definition

$$A^{(0)} = R \quad \text{and} \quad A^{(1)} = \text{ideal generated by the components of } A.$$

Furthermore

$$I_r(A) = 0 \quad \text{for} \quad r > \min(p, q)$$

and the inclusions

$$R = I_0(A) \supset I_1(A) \supset I_2(A) \supset \cdots$$

By Proposition 10.1, we also have

(1) $$I_r(AB) \subset I_r(A) \cap I_r(B).$$

Therefore, if $A = UBU'$ where U, U' are square matrices of determinant 1, then

(2) $$I_r(A) = I_r(B).$$

Next, let E be an R-module. Let $x_1, \dots, x_q$ be generators of E. Then we may form the matrix of relations $(a_1, \dots, a_q) \in R^q$ such that

$$\sum_{i=1}^{q} a_i x_i = 0.$$

Suppose first we take only finitely many relations, thus giving rise to a $p \times q$ matrix A. We form the determinant ideal $I_r(A)$. We let the **determinant ideals** of the family of generators be:

$$I_r(x_1, \dots, x_q) = I_r(x) = \text{ideal generated by } I_r(A) \text{ for all } A.$$

Thus we may in fact take the infinite matrix of relations, and say that $I_r(x)$ is generated by the determinants of all $r \times r$ submatrices. The inclusion relations of (1) show that

$$R = I_0(x) \supset I_1(x) \supset I_2(x) \supset \cdots$$
$$I_r(x) = 0 \quad \text{if} \quad r > q.$$

Furthermore, it is easy to see that if we form a submatrix M of the matrix of all relations by taking only a family of relations which generate the ideal of all relations in R^q, then we have

$$I_r(M) = I_r(x).$$

We leave the verification to the reader. We can take M to be a finite matrix when E is finitely presented, which happens if R is Noetherian.

In terms of this representation of a module as a quotient of R^q, we get the following characterization.

Proposition 2.2. *Let $R^q \to E \to 0$ be a representation of E as a quotient of R^q, and let $x_1, \dots, x_q$ be the images of the unit vectors in R^q. Then $I_r(x)$ is the ideal generated by all values*

$$\lambda(w_1, \dots, w_r)$$

where $w_1, \dots, w_r \in \operatorname{Ker}(R^q \to E)$ and $\lambda \in L_a^r(R^q, R)$.

Proof. This is immediate from the definition of the determinant ideal.

The above proposition can be useful to replace a matrix computation by a more conceptual argument with fewer indices. The reader can profitably translate some of the following matrix arguments in these more invariant terms.

We now change the numbering, and let the **Fitting ideals** be:

$$F_k(x) = I_{q-k}(x) \quad \text{for} \quad 0 \leqq k \leqq q$$

$$F_k(x) = R \qquad \text{when} \quad k > q.$$

Lemma 2.3. *The Fitting ideal $F_k(x)$ does not depend on the choice of generators (x).*

Proof. Let $y_1, \ldots, y_s$ be elements of E. We shall prove that

$$I_r(x) = I_{r+s}(x, y).$$

The relations of (x, y) constitute a matrix of the form

$$W = \begin{pmatrix} a_{11} & \cdots & a_{1q} & 0 & & \cdots & 0 \\ \vdots & & \vdots & \vdots & & & \vdots \\ a_{p1} & \cdots & a_{pq} & 0 & & \cdots & 0 \\ b_{11} & \cdots & b_{1q} & 1 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & & \vdots \\ b_{s1} & \cdots & b_{sq} & 0 & & \cdots & 1 \end{pmatrix}$$

By elementary column operations, we can change this to a matrix

$$\begin{pmatrix} A & 0 \\ 0 & 1_s \end{pmatrix}$$

and such operations do not change the determinant ideals by (2). Then we conclude that for all $r \geqq 0$ we have

$$I_r(A) = I_{r+s}(W) \subset I_{r+s}(x, y).$$

This proves that $I_r(x) \subset I_{r+s}(x, y)$.

Conversely, let C be a matrix of relations between the generators (x, y). We also have a matrix of relations

$$Z = \begin{pmatrix} & & C & & & \\ b_{11} & \cdots & b_{1q} & 1 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ b_{s1} & \cdots & b_{sq} & 0 & \cdots & 1 \end{pmatrix}$$

By elementary row operations, we can bring this matrix into the same shape

as B above, with some matrix of relations A' for (x), namely

$$Z' = \begin{pmatrix} A' & 0 \\ B & 1_s \end{pmatrix}$$

Then

$$I_r(A') = I_{r+s}(Z') = I_{r+s}(Z) \supset I_{r+s}(C),$$

whence $I_{r+s}(C) \subset I_r(x)$. Taking all possible matrices of relations C shows that $I_{r+s}(x, y) \subset I_r(x)$, which combined with the previous inequality yields $I_{r+s}(x, y) = I_r(x)$.

Now given two families of generators (x) and (y), we simply put them side by side (x, y) and use the new numbering for the F_k to conclude the proof of the lemma.

Now let E be a finitely generated R-module with presentation

$$0 \to K \to R^q \to E \to 0,$$

where the sequence is exact and K is defined as the kernel. Then K is generated by q-vectors, and can be viewed as an infinite matrix. The images of the unit vectors in R^q are generators $(x_1, \ldots, x_q)$. We define the **Fitting ideal** of the module to be

$$F_k(E) = F_k(x).$$

Lemma 2.3 shows that the ideal is independent of the choice of presentation. The inclusion relations of a determinant ideal $I_r(A)$ of a matrix now translate into reverse inclusion relations for the Fitting ideals, namely:

Proposition 2.4.

(i) *We have*

$$F_0(E) \subset F_1(E) \subset F_2(E) \subset \cdots$$

(ii) *If E can be generated by q elements, then*

$$F_q(E) = R.$$

(iii) *If E is finitely presented then $F_k(E)$ is finitely generated for all k.*

This last statement merely repeats the property that the determinant ideals of a matrix can be generated by the determinants associated with a finite submatrix if the row space of the matrix is finitely generated.

Example. Let $E = R^q$ be the free module of dimension q. Then:

$$F_k(E) = \begin{cases} 0 & \text{if } 0 \leqq k < q \\ R & \text{if } k \geqq q. \end{cases}$$

This is immediate from the definitions and the fact that the only relation of a basis for E is the trivial one.

The Fitting ideal $F_0(E)$ is called the **zero-th** or **initial Fitting ideal**. In some applications it is the only one which comes up, in which case it is called "**the**" **Fitting ideal** $F(E)$ of E. It is the ideal generated by all $q \times q$ determinants in the matrix of relations of q generators of the module.

For any module E we let $\text{ann}_R(E)$ be the annihilator of E in R, that is the set of elements $a \in R$ such that $aE = 0$.

Proposition 2.5. *Suppose that E can be generated by q elements. Then*

$$(\text{ann}_R(E))^q \subset F(E) \subset \text{ann}_R(E).$$

In particular, if E can be generated by one element, then

$$F(E) = \text{ann}_R(E).$$

Proof. Let $x_1, \ldots, x_q$ be generators of E. Let $a_1, \ldots, a_q$ be elements of R annihilating E. Then the diagonal matrix whose diagonal components are $a_1, \ldots, a_q$ is a matrix of relations, so the definition of the Fitting ideal shows that the determinant of this matrix, which is the product $a_1 \cdots a_q$ lies in $I_q(E) \subset F_0(E)$. This proves the inclusion

$$\text{ann}_R(E)^q \subset F(E).$$

Conversely, let A be a $q \times q$ matrix of relations between $x_1, \ldots, x_q$. Then $\det(A)x_i = 0$ for all i so $\det(A) \in \text{ann}_R(E)$. Since $F(E)$ is generated by such determinants, we get the reverse inclusion which proves the proposition.

Corollary 2.6. *Let $E = R/\mathfrak{a}$ for some ideal $\mathfrak{a}$. Then $F(E) = \mathfrak{a}$.*

Proof. The module $R/\mathfrak{a}$ can be generated by one element so the corollary is an immediate consequence of the proposition.

Proposition 2.7. *Let*

$$0 \to E' \to E \to E'' \to 0$$

be an exact sequence of finite R-modules. For integers $m, n \geqq 0$ we have

$$F_m(E')F_n(E'') \subset F_{m+n}(E).$$

In particular for $F = F_0$,

$$F(E')F(E'') \subset F(E).$$

Proof. We may assume E' is a submodule of E. We pick generators $x_1, \ldots, x_p$ of E' and elements $y_1, \ldots, y_q$ in E such that their images $y_1'', \ldots, y_q''$ in E'' generate E''. Then (x, y) is a family of generators for E. Suppose first that $m \leqq p$ and $n \leqq q$. Let A be a matrix of relations among $y_1'', \ldots, y_q''$ with q columns. If $(a_1, \ldots, a_q)$ is such a relation, then

$$a_1 y_1 + \cdots + a_q y_q \in E'$$

so there exist elements $b_1, \ldots, b_p \in R$ such that

$$\sum a_i y_i + \sum b_j x_j = 0.$$

Thus we can find a matrix B with p columns and the same number of rows as A such that (B, A) is a matrix of relations of (x, y). Let C be a matrix of relations of $(x_1, \ldots, x_p)$. Then

$$\begin{pmatrix} B & A \\ C & 0 \end{pmatrix}$$

is a matrix of relations of (x, y). If D'' is a $(q - n) \times (q - n)$ subdeterminant of A and D' is a $(p - m) \times (p - m)$ subdeterminant of C then $D''D'$ is a

$$(p + q - m - n) \times (p + q - m - n)$$

subdeterminant of the matrix

$$\begin{pmatrix} B & A \\ C & 0 \end{pmatrix}$$

and $D''D' \in F_{m+n}(E)$. Since $F_m(E')$ is generated by determinants like D' and $F_n(E'')$ is generated by determinants like D'', this proves the proposition in the present case.

If $m > p$ and $n > q$ then $F_{m+n}(E) = F_m(E') = F_n(E'') = R$ so the proposition is trivial in this case.

Say $m \leqq p$ and $n > q$. Then $F_n(E'') = R = F_q(E'')$ and hence

$$F_m(E')F_n(E'') = F_q(E'')F_m(E') \subset F_{p+n}(E) \subset F_{m+n}(E)$$

where the inclusion follows from the first case. A similar argument proves the remaining case with $m > p$ and $n \leqq q$. This concludes the proof.

Proposition 2.8. *Let E', E'' be finite R-modules. For any integer $n \geqq 0$ we have*

$$F_n(E' \oplus E'') = \sum_{r+s=n} F_r(E')F_s(E'').$$

Proof. Let $x_1, \ldots, x_p$ generate E^i and $y_1, \ldots, y_q$ generate E''. Then (x, y) generate $E' \oplus E''$. By Proposition 2.6 we know the inclusion

$$\sum F_r(E')F_s(E'') \subset F_n(E' \oplus E''),$$

so we have to prove the converse. If $n \geqq p + q$ then we can take $r \geqq p$ and $s \geqq q$ in which case

$$F_r(E') = F_s(E'') = F_n(E) = R$$

and we are done. So we assume $n < p + q$. A relation between (x, y) in the direct sum splits into a relation for (x) and a relation for (y). The matrix of relations for (x, y) is therefore of the form

$$C = \begin{pmatrix} A' & 0 \\ 0 & A'' \end{pmatrix}$$

where A' is the matrix of relations for (x) and A'' the matrix of relations for (y). Thus

$$F_n(E' \oplus E'') = \sum_C I_{p+q-n}(C)$$

where the sum is taken over all matrices C as above. Let D be a

$$(p + q - n) \times (p + q - n)$$

subdeterminant. Then D has the form

$$D = \begin{vmatrix} B' & 0 \\ 0 & B'' \end{vmatrix}$$

where B' is a $k' \times (p - r)$ matrix, and B'' is a $k'' \times (q - s)$ matrix with some positive integers k', k'', r, s satisfying

$$k' + k'' = p + q - n \quad \text{and} \quad r + s = n.$$

Then $D = 0$ unless $k' = p - r$ and $k'' = q - s$. In that case

$$D = \det(B')\det(B'') \in F_r(E')F_s(E''),$$

which proves the reverse inclusion and concludes the proof of the proposition.

Corollary 2.9. *Let*

$$E = \bigoplus_{i=1}^{s} R/\mathfrak{a}_i$$

where $\mathfrak{a}_i$ is an ideal. Then $F(E) = \mathfrak{a}_1 \cdots \mathfrak{a}_s$.

Proof. This is really a corollary of Proposition 2.8 and Corollary 2.6.

§3. UNIVERSAL DERIVATIONS AND THE DE RHAM COMPLEX

In this section, all rings R, A, etc. are assumed commutative.

Let A be an R-algebra and M an A-module. By a **derivation** $D: A \to M$ (over R) we mean an R-linear map satisfying the usual rules

$$D(ab) = aDb + bDa.$$

Note that $D(1) = 2D(1)$ so $D(1) = 0$, whence $D(R) = 0$. Such derivations form an A-module $\mathrm{Der}_R(A, M)$ in a natural way, where aD is defined by $(aD)(b) = aDb$.

By a **universal derivation** for A over R, we mean an A-module Ω, and a derivation

$$d : A \to \Omega$$

such that, given a derivation $D : A \to M$ there exists a unique A-homomorphism $f : \Omega \to M$ making the following diagram commutative:

$$\begin{array}{ccc} A & \xrightarrow{d} & \Omega \\ & {}_D\searrow \quad \swarrow_f & \\ & M & \end{array}$$

It is immediate from the definition that a universal derivation (d, Ω) is uniquely determined up to a unique isomorphism. By definition, we have a functorial isomorphism

$$\boxed{\mathrm{Der}_R(A, M) \approx \mathrm{Hom}_A(\Omega, M).}$$

We shall now prove the existence of a universal derivation.

The following general remark will be useful. Let

$$f_1, f_2 : A \to B$$

be two homomorphisms of R-algebras, and let J be an ideal in B such that $J^2 = 0$. Assume that $f_1 \equiv f_2 \bmod J$; this means that $f_1(x) \equiv f_2(x) \bmod J$ for all x in A. Then

$$D = f_2 - f_1$$

is a derivation. This fact is immediately verified as follows:

$$\begin{aligned} f_2(ab) = f_2(a)f_2(b) &= [f_1(a) + D(a)][f_1(b) + D(b)] \\ &= f_1(ab) + f_1(b)D(a) + f_1(a)D(b). \end{aligned}$$

But the A-module structure of J is given via f_1 or f_2 (which amount to the same thing in light of our assumptions on f_1, f_2), so the fact is proved.

Let the tensor product be taken over R.

Let $\mathbf{m}_A : A \otimes A \to A$ be the multiplication homomorphism, such that $\mathbf{m}_A(a \otimes b) = ab$. Let $J = \operatorname{Ker} \mathbf{m}_A$. We define the module of **differentials**

$$\Omega_{A/R} = J/J^2,$$

as an ideal in $(A \otimes A)/J^2$. The A-module structure will always be given via the embedding on the first factor:

$$A \to A \otimes A \quad \text{by} \quad a \mapsto a \otimes 1.$$

Note that we have a direct sum decomposition of A-modules

$$A \otimes A = (A \otimes 1) \oplus J,$$

and therefore

$$(A \otimes A)/J^2 = (A \otimes 1) \oplus J/J^2.$$

Let

$$d : A \to J/J^2 \text{ be the } R\text{-linear map } a \mapsto 1 \otimes a - a \otimes 1 \bmod J^2.$$

Taking $f_1 : a \mapsto a \otimes 1$ and $f_2 : a \mapsto 1 \otimes a$, we see that $d = f_2 - f_1$. Hence d is a derivation when viewed as a map into J/J^2.

We note that J is generated by elements of the form

$$\sum x_i \, dy_i.$$

Indeed, if $\sum x_i \otimes y_i \in J$, then by definition $\sum x_i y_i = 0$, and hence

$$\sum x_i \otimes y_i = \sum x_i(1 \otimes y_i - y_i \otimes 1),$$

according to the A-module structure we have put on $A \otimes A$ (operation of A on the left factor.)

Theorem 3.1. *The pair $(J/J^2, d)$ is universal for derivations of A. This means: Given a derivation D: $A \to M$ there exists a unique A-linear map $f : J/J^2 \to M$ making the following diagram commutative.*

$$\begin{array}{ccc} A & \xrightarrow{d} & J/J^2 \\ & {}_D\searrow \quad \swarrow_f & \\ & M & \end{array}$$

Proof. There is a unique R-bilinear map

$$f: A \otimes A \to M \quad \text{given by} \quad x \otimes y \mapsto xDy,$$

which is A-linear by our definition of the A-module structure on $A \otimes A$. Then by definition, the diagram is commutative on elements of A, when we take f restricted to J, because

$$f(1 \otimes y - y \otimes 1) = Dy.$$

Since J/J^2 is generated by elements of the form $x\, dy$, the uniqueness of the map in the diagram of the theorem is clear. This proves the desired universal property.

We may write the result expressed in the theorem as a formula

$$\boxed{\mathrm{Der}_R(A, M) \approx \mathrm{Hom}_A(J/J^2, M).}$$

The reader will find exercises on derivations which give an alternative way of constructing the universal derivation, especially useful when dealing with finitely generated algebras, which are factors of polynomial rings.

I insert here without proofs some further fundamental constructions, important in differential and algebraic geometry. The proofs are easy, and provide nice exercises.

Let $R \to A$ be an R-algebra of commutative rings. For $i \geqq 0$ define

$$\Omega^i_{A/R} = \bigwedge^i \Omega^1_{A/R},$$

where $\Omega^0_{A/R} = A$.

Theorem 3.2. *There exists a unique sequence of R-homomorphisms*

$$d_i : \Omega^i_{A/R} \to \Omega^{i+1}_{A/R}$$

such that for $\omega \in \Omega^i$ and $\eta \in \Omega^j$ we have

$$d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^i \omega \wedge d\eta.$$

Furthermore $d \circ d = 0$.

The proof will be left as an exercise.

Recall that a **complex** of modules is a sequence of homomorphisms

$$\cdots \to E^{i-1} \xrightarrow{d^{i-1}} E^i \xrightarrow{d^i} E^{i+1} \to$$

such that $d^i \circ d^{i-1} = 0$. One usually omits the superscript on the maps d. With this terminology, we see that the $\Omega^i_{A/R}$ form a complex, called the **De Rham complex**.

Theorem 3.3. *Let k be a field of characteristic 0, and let $A = k[X_1, \ldots, X_n]$ be the polynomial ring in n variables. Then the De Rham complex*

$$0 \to k \to A \to \Omega^1_{A/k} \to \cdots \to \Omega^n_{A/k} \to 0$$

is exact.

Again the proof will be left as an exercise. *Hint*: Use induction and integrate formally.

Other results concerning connections will be found in the exercises below.

§4. THE CLIFFORD ALGEBRA

Let k be a field. By an **algebra** throughout this section, we mean a k-algebra given by a ring homomorphism $k \to A$ such that the image of k is in the center of A.

Let E be a finite dimensional vector space over the field k, and let g be a symmetric form on E. We would like to find a universal algebra over k, in which we can embed E, and such that the square in the algebra corresponds to the value of the quadratic form in E. More precisely, by a **Clifford algebra** for g, we shall mean a k-algebra $C(g)$, also denoted by $C_g(E)$, and a linear map $\rho: E \to C(g)$ having the following property: If $\psi : E \to L$ is a linear map of E into a k-algebra L such that

$$\psi(x)^2 = g(x, x) \cdot 1 \qquad (1 = \text{unit element of } L)$$

for all $x \in E$, then there exists a unique algebra-homomorphism

$$C(\psi) = \psi_* : C(g) \to L$$

such that the following diagram is commutative:

$$\begin{array}{ccc} E & \xrightarrow{\rho} & C(g) \\ & {\scriptstyle\psi}\searrow \quad \swarrow{\scriptstyle\psi_*} & \\ & L & \end{array}$$

By abstract nonsense, a Clifford algebra for g is uniquely determined, up to a unique isomorphism. Furthermore, it is clear that if $(C(g), \rho)$ exists, then $C(g)$ is generated by the image of ρ, i.e. by $\rho(E)$, *as an algebra over k.*

We shall write $\rho = \rho_g$ if it is necessary to specify the reference to g explicitly.

We have trivially

$$\rho(x)^2 = g(x, x) \cdot 1$$

for all $x \in E$, and

$$\rho(x)\rho(y) + \rho(y)\rho(x) = 2g(x, y) \cdot 1$$

as one sees by replacing x by $x + y$ in the preceding relation.

Theorem 4.1. *Let g be a symmetric bilinear form on a finite dimensional vector space E over k. Then the Clifford algebra $(C(g), \rho)$ exists. The map ρ in injective, and $C(g)$ has dimension 2^n over k, if $n = \dim E$.*

Proof. Let $T(E)$ be the tensor algebra as in Chapter XVI, §7. In that algebra, we let I_g be the two-sided ideal generated by all elements

$$x \otimes x - g(x, x) \cdot 1 \text{ for } x \in E.$$

We define $C_g(E) = T(E)/I_g$. Observe that E is naturally embedded in $T(E)$ since

$$T(E) = k \oplus E \oplus (E \otimes E) \oplus \cdots .$$

Then the natural embedding of E in TE followed by the canonical homomorphisms of $T(E)$ onto $C_g(E)$ defines our k-linear map $\rho : E \to C_g(E)$. It is immediate from the universal property of the tensor product that $C_g(E)$ as just defined satisfies the universal property of a Clifford algebra, which therefore exists. The only problem is to prove that it has the stated dimension over k.

We first prove that the dimension is $\leqq 2^n$. We give a proof only when the characteristic of k is $\neq 2$ and leave characteristic 2 to the reader. Let $\{v_1, \ldots, v_n\}$ be an orthogonal basis of E as given by Theorem 3.1 of Chapter XV. Let $e_i = \psi(v_i)$, where $\psi : E \to L$ is given as in the beginning of the section. Let $c_i = g(v_i, v_i)$. Then we have the relations

$$e_i^2 = c_i, \qquad e_i e_j = -e_j e_i \text{ for all } i \neq j.$$

This immediately implies that the subalgebra of L generated by $\psi(E)$ over k is generated as a vector space over k by all elements

$$e_1^{\nu_1} \cdots e_n^{\nu_n} \text{ with } \nu_i = 0 \text{ or } 1 \text{ for } i = 1, \ldots, n.$$

Hence the dimension of this subalgebra is $\leqq 2^n$. In particular, $\dim C_g(E) \leqq 2^n$ as desired.

There remains to show that there exists at least one $\psi: E \to L$ such that L is generated by $\psi(E)$ as an algebra over k, and has dimension 2^n; for in that case, the homomorphism $\psi_* : C_g(E) \to L$ being surjective, it follows that dim $C_g(E) \geqq 2^n$ and the theorem will be proved. We construct L in the following way. We first need some general notions.

Let M be a module over a commutative ring. Let $i, j \in \mathbf{Z}/2\mathbf{Z}$. Suppose M is a direct sum $M = M_0 \oplus M_1$ where 0, 1 are viewed as the elements of $\mathbf{Z}/2\mathbf{Z}$. We then say that M is **$\mathbf{Z}/2\mathbf{Z}$-graded**. If M is an algebra over the ring, we say

it is a **$\mathbf{Z}/2\mathbf{Z}$-graded algebra** if $M_i M_j \subset M_{i+j}$ for all $i, j \in \mathbf{Z}/2\mathbf{Z}$. We simply say **graded**, omitting the $\mathbf{Z}/2\mathbf{Z}$ prefix when the reference to $\mathbf{Z}/2\mathbf{Z}$ is fixed throughout a discussion, which will be the case in the rest of this section.

Let A, B be graded modules as above, with $A = A_0 \oplus A_1$ and $B = B_0 \oplus B_1$. Then the tensor product $A \otimes B$ has a direct sum decomposition

$$A \otimes B = \bigoplus_{i,j} A_i \otimes B_j.$$

We define a grading on $A \otimes B$ by letting $(A \otimes B)_0$ consist of the sum over indices i, j such that $i + j = 0$ (in $\mathbf{Z}/2\mathbf{Z}$), and $(A \otimes B)_1$ consist of the sum over the indices i, j such that $i + j = 1$.

Suppose that A, B are graded algebras over the given commutative ring. There is a unique bilinear map of $A \otimes B$ into itself such that

$$(a \otimes b)(a' \otimes b') = (-1)^{ij} aa' \otimes bb'$$

if $a' \in A_i$ and $b \in B_j$. Just as in Chapter XVI, §6, one verifies associativity and the fact that this product gives rise to a graded algebra, whose product is called the **super tensor product**, or **super product**. As a matter of notation, when we take the super tensor product of A and B, we shall denote the resulting algebra by

$$A \otimes_{su} B$$

to distinguish it from the ordinary algebra $A \otimes B$ of Chapter XVI, §6.

Next suppose that E has dimension 1 over k. Then the factor polynomial ring $k[X]/(x^2 - c_1)$ is immediately verified to be the Clifford algebra in this case. We let t_1 be the image of X in the factor ring, so $C_g(E) = k[t_1]$ with $t_1^2 = c_1$. The vector space E is imbedded as kt_1 in the direct sum $k \oplus kt_1$.

In general we now take the super tensor product inductively:

$$C_g(E) = k[t_1] \otimes_{su} k[t_2] \otimes_{su} \cdots \otimes_{su} k[t_n], \text{ with } k[t_i] = k[X]/(X^2 - c_i).$$

Its dimension is 2^n. Then E is embedded in $C_g(E)$ by the map

$$a_1 v_1 + \cdots + a_n v_n \mapsto a_1 t_1 \oplus \cdots \oplus a_n t_n.$$

The desired commutation rules among t_i, t_j are immediately verified from the definition of the super product, thus concluding the proof of the dimension of the Clifford algebra.

Note that the proof gives an explicit representation of the relations of the algebra, which also makes it easy to compute in the algebra. Note further that the alternating algebra of a free module is a special case, taking $c_i = 0$ for all i. Taking the c_i to be algebraically independent shows that the alternating algebra is a specialization of the generic Clifford algebra, or that Clifford algebras are what one calls perturbations of the alternating algebra. Just as for the alternating algebra, we have immediately from the construction:

Theorem 4.2. *Let g, g' by symmetric forms on E, E' respectively. Then we*

have an algebra isomorphism

$$C(g \oplus g') \approx C(g) \otimes_{su} C(g').$$

Examples. Clifford algebras have had increasingly wide applications in physics, differential geometry, topology, group representations (finite groups and Lie groups), and number theory. First, in topology I refer to Adams [Ad 62] and [ABS 64] giving applications of the Clifford algebra to various problems in topology, notably a description of the way Clifford algebras over the reals are related to the existence of vector fields on spheres. The multiplication in the Clifford algebra gives rise to a multiplication on the sphere, whence to vector fields. [ABS 64] also gives a number of computations related to the Clifford algebra and its applications to topology and physics. For instance, let $E = \mathbf{R}^n$ and let g be the negative of the standard dot product. Or more invariantly, take for E an n-dimensional vector space over $\mathbf{R}$, and let g be a *negative definite* symmetric form on E. Let $C_n = C(g)$.

The operation

$$v_1 \otimes \cdots \otimes v_r \mapsto v_r \otimes \cdots \otimes v_1 = (v_1 \otimes \cdots \otimes v_r)^* \text{ for } v_i \in E$$

induces an endomorphism of $T^r(E)$ for $r \geqq 0$. Since $v \otimes v - g(v, v) \cdot 1$ (for $v \in E$) is invariant under this operation, there is an induced endomorphism $* : C_n \to C_n$, which is actually an involution, that is $x^{**} = x$ and $(xy)^* = y^*x^*$ for $x \in C_n$. We let Spin(n) be the subgroup of units in C_n generated by the unit sphere in E (i.e. the set of elements such that $g(v, v) = -1$), and lying in the even part of C_n. Equivalently, Spin(n) is the group of elements x such that $xx^* = 1$. The name dates back to Dirac who used this group in his study of electron spin. Topologists and others view that group as being the universal covering group of the special orthogonal group $SO(n) = SU_n(\mathbf{R})$.

An account of some of the results of [Ad 62] and [ABS 64] will also be found in [Hu 75], Chapter 11. Second I refer to two works encompassing two decades, concerning the heat kernel, Dirac operator, index theorem, and number theory, ranging from Atiyah, Bott and Patodi [ABP 73] to Faltings [Fa 91], see especially §4, entitled "The local index theorem for Dirac operators". The vector space to which the general theory is applied is mostly the cotangent space at a point on a manifold. I recommend the book [BGV 92], Chapter 3.

Finally, I refer to Bröcker and Tom Dieck for applications of the Clifford algebra to representation theory, starting with their Chapter I, §6, [BtD 85].

Bibliography

[Ad 62] F. ADAMS, Vector Fields on Spheres, *Ann. Math.* **75** (1962) pp. 603–632

[ABP 73] M. ATIYAH, R. BOTT, V. PATODI, On the heat equation and the index theorem, *Invent. Math.* **19** (1973) pp. 270–330; erratum **38** (1975) pp. 277–280

[ABS 64] M. ATIYAH, R. BOTT, A. SHAPIRO, Clifford Modules, *Topology* **Vol. 3, Supp. 1** (1964) pp. 3–38

[BGV 92] N. BERLINE, E. GETZLER, and M. VERGNE, *Heat Kernels and Dirac Operators*, Springer Verlag, 1992

[BtD 85] T. BRÖCKER and T. TOM DIECK, *Representations of Compact Lie Groups*, Springer Verlag 1985

[Fa 91] G. FALTINGS, *Lectures on the arithmetic Riemann-Roch theorem*, Annals of Math. Studies 1991

[Hu 75] D. HUSEMOLLER, *Fibre Bundles*, Springer Verlag, Second Edition, 1975

EXERCISES

1. Let E be a finite dimensional vector space over a field k. Let $x_1, \dots, x_p$ be elements of E such that $x_1 \wedge \cdots \wedge x_p \neq 0$, and similarly $y_1 \wedge \cdots \wedge y_p \neq 0$. If $c \in k$ and

$$x_1 \wedge \cdots \wedge x_p = cy_1 \wedge \cdots \wedge y_p$$

show that $x_1, \dots, x_p$ and $y_1, \dots, y_p$ generate the same subspace. Thus non-zero decomposable vectors in $\bigwedge^p E$ up to non-zero scalar multiples correspond to p-dimensional subspaces of E.

2. Let E be a free module of dimension n over the commutative ring R. Let $f: E \to E$ be a linear map. Let $\alpha_r(f) = \operatorname{tr} \bigwedge^r(f)$, where $\bigwedge^r(f)$ is the endomorphism of $\bigwedge^r(E)$ into itself induced by f. We have

$$\alpha_0(f) = 1, \qquad \alpha_1(f) = \operatorname{tr}(f), \qquad \alpha_n(f) = \det f,$$

and $\alpha_r(f) = 0$ if $r > n$. Show that

$$\det(1 + f) = \sum_{r \geqq 0} \alpha_r(f).$$

[*Hint*: As usual, prove the statement when f is represented by a matrix with variable coefficients over the integers.] Interpret the $\alpha_r(f)$ in terms of the coefficients of the characteristic polynomial of f.

3. Let E be a finite dimensional free module over the commutative ring R. Let $E^\vee$ be its dual module. For each integer $r \geqq 1$ show that $\bigwedge^r E$ and $\bigwedge^r E^\vee$ are dual modules to each other, under the bilinear map such that

$$(v_1 \wedge \cdots \wedge v_r, v_1' \wedge \cdots \wedge v_r') \mapsto \det(\langle v_i, v_j' \rangle)$$

where $\langle v_i, v_j' \rangle$ is the value of v_j' on v_i, as usual, for $v_i \in E$ and $v_j' \in E^\vee$.

4. Notation being as in the preceding exercise, let F be another R-module which is free, finite dimensional. Let $f: E \to F$ be a linear map. Relative to the bilinear map of the preceding exercise, show that the transpose of $\bigwedge^r f$ is $\bigwedge^r({}^t f)$, i.e. is equal to the r-th alternating product of the transpose of f.

5. Let R be a commutative ring. If E is an R-module, denote by $L_a^r(E)$ the module of

r-multilinear alternating maps of E into R itself (i.e. the r-multilinear alternating forms on E). Let $L_a^0(E) = R$, and let

$$\Omega(E) = \bigoplus_{r=0}^{\infty} L_a^r(E).$$

Show that $\Omega(E)$ is a graded R-algebra, the multiplication being defined as follows. If $\omega \in L_a^r(E)$ and $\psi \in L_a^s(E)$, and $v_1, \ldots, v_{r+s}$ are elements of E, then

$$(\omega \wedge \psi)(v_1, \ldots, v_{r+s}) = \sum \epsilon(\sigma)\omega(v_{\sigma 1}, \ldots, v_{\sigma r})\psi(v_{\sigma(r+1)}, \ldots, v_{\sigma s}),$$

the sum being taken over all permutations σ of $(1, \ldots, r+s)$ such that $\sigma 1 < \cdots < \sigma r$ and $\sigma(r+1) < \cdots < \sigma s$.

Derivations

In the following exercises on derivations, all rings are assumed commutative. Among other things, the exercises give another proof of the existence of universal derivations.

Let $R \to A$ be a R-algebra (of commutative rings, according to our convention). We denote the module of universal derivations of A over R by $(d_{A/R}, \Omega_{A/R}^1)$, but we do not assume that it necessarily exists. Sometimes we write d instead of $d_{A/R}$ for simplicity if the reference to A/R is clear.

6. Let $A = R[X_\alpha]$ be a polynomial ring in variables X_α, where α ranges over some indexing set, possibly infinite. Let Ω be the free A-module on the symbols dX_α, and let

$$d : A \to \Omega$$

be the mapping defined by

$$df(X) = \sum_\alpha \frac{\partial f}{\partial X_\alpha} dX_\alpha.$$

Show that the pair (d, Ω) is a universal derivation $(d_{A/R}, \Omega_{A/R}^1)$.

7. Let $A \to B$ be a homomorphism of R-algebras. Assume that the universal derivations for A/R, B/R, and B/A exist. Show that one has a natural exact sequence:

$$B \otimes_A \Omega_{A/R}^1 \to \Omega_{B/R}^1 \to \Omega_{B/A}^1 \to 0.$$

[*Hint*: Consider the sequence

$$0 \to \mathrm{Der}_A(B, M) \to \mathrm{Der}_R(B, M) \to \mathrm{Der}_R(A, M)$$

which you prove is exact. Use the fact that a sequence of B-modules

$$N' \to N \to N'' \to 0$$

is exact if and only if its Hom into M is exact for every B-module M. Apply this to the sequence of derivations.]

8. Let $R \to A$ be an R-algebra, and let I be an ideal of A. Let $B = A/I$. Suppose that the universal derivation of A over R exists. Show that the universal derivation of B over R

also exists, and that there is a natural exact sequence

$$I/I^2 \xrightarrow{d_{A/R}} B \otimes_A \Omega^1_{A/R} \to \Omega^1_{B/R} \to 0.$$

[*Hint*: Let M be a B-module. Show that the sequence

$$0 \to \mathrm{Der}_R(B, M) \to \mathrm{Der}_R(A, M) \to \mathrm{Hom}_B(I/I^2, M)$$

is exact.]

9. Let $R \to B$ be an R-algebra. Show that the universal derivation of B over R exists as follows. Represent B as a quotient of a polynomial ring, possibly in infinitely many variables. Apply Exercises 6 and 7.

10. Let $R \to A$ be an R-algebra. Let S_0 be a multiplicative subset of R, and S a multiplicative subset of A such that S_0 maps into S. Show that the universal derivation of $S^{-1}A$ over $S_0^{-1}R$ is $(d, S^{-1}\Omega^1_{A/R})$, where

$$d(a/s) = (sd_{A/R}(a) - ad_{A/R}(s))/s^2.$$

11. Let B be an R-algebra and M a B-module. On $B \oplus M$ define a product

$$(b, x)(b', y) = (bb', by + b'x).$$

Show that $B \oplus M$ is a B-algebra, if we identify an element $b \in B$ with $(b, 0)$. For any R-algebra A, show that the algebra homomorphisms $\mathrm{Hom}_{\mathrm{Alg}/R}(A, B \oplus M)$ consist of pairs (φ, D), where $\varphi : A \to B$ is an algebra homomorphism, and $D : A \to M$ is a derivation for the A-module structure on M induced by φ.

12. Let A be an R-algebra. Let $\varepsilon : A \to R$ be an algebra homomorphism, which we call an **augmentation**. Let M be an R-module. Define an A-module structure on M via ε, by

$$a \cdot x = \varepsilon(a)x \quad \text{for} \quad a \in A \quad \text{and} \quad x \in M.$$

Write M_ε to denote M with this new module structure. Let:

$$\mathrm{Der}_\varepsilon(A, M) = A\text{-module of derivations for the } \varepsilon\text{-module structure on } M$$

$$I = \mathrm{Ker}\,\varepsilon.$$

Then $\mathrm{Der}_\varepsilon(A, M)$ is an A/I-module. Note that there is an R-module direct sum decomposition $A = R \oplus I$. Show that there is a natural A-module isomorphism

$$\Omega_{A/R}/I\Omega_{A/R} \approx I/I^2$$

and an R-module isomorphism

$$\mathrm{Der}_\varepsilon(A, M) \approx \mathrm{Hom}_R(I/I^2, M).$$

In particular, let $\eta : A \to I/I^2$ be the projection of A on I/I^2 relative to the direct sum decomposition $A = R \oplus I$. Then η is the universal ε-derivation.

Derivations and connections

13. Let $R \to A$ be a homomorphism of commutative rings, so we view A as an R-algebra.

Let E be an A-module. A **connection** on E is a homomorphism of abelian groups

$$\nabla : E \to \Omega^1_{A/R} \otimes_A E$$

such that for $a \in A$ and $x \in E$ we have

$$\nabla(ax) = a\nabla(x) + da \otimes x,$$

where the tensor product is taken over A unless otherwise specified. The kernel of ∇, denoted by E_∇, is called the **submodule of horizontal elements**, or the **horizontal submodule** of (E, ∇).

(a) For any integer $i \geqq 1$, define

$$\Omega^i_{A/R} = \bigwedge^i \Omega^1_{A/R}.$$

Show that ∇ can be extended to a homomorphism of R-modules

$$\nabla_i : \Omega^i_{A/R} \otimes E \to \Omega^{i+1}_{A/R} \otimes E$$

by

$$\nabla_i(\omega \otimes x) = d\omega \otimes x + (-1)^i \omega \wedge \nabla(x).$$

(b) Define the **curvature** of the connection to be the map

$$K = \nabla_1 \circ \nabla : E \to \Omega^2_{A/R} \otimes_A E.$$

Show that K is an A-homomorphism. Show that

$$\nabla_{i+1} \circ \nabla_i(\omega \otimes x) = \omega \wedge K(x)$$

for $\omega \in \Omega^i_{A/R}$ and $x \in E$.

(c) Let $\mathrm{Der}(A/R)$ denote the A-module of derivations of A into itself, over R. Let ∇ be a connection on E. Show that ∇ induces a unique A-linear map

$$\nabla : \mathrm{Der}(A/R) \to \mathrm{End}_R(E)$$

such that

$$\nabla(D)(ax) = D(a)x + a\nabla(D)(x).$$

(d) Prove the formula

$$[\nabla(D_1), \nabla(D_2)] - \nabla([D_1, D_2]) = (D_1 \wedge D_2)(K).$$

In this formula, the bracket is defined by $[f, g] = f \circ g - g \circ f$ for two endomorphisms f, g of E. Furthermore, the right-hand side is the composed mapping

$$E \xrightarrow{K} \Omega^2_{A/R} \otimes E \xrightarrow{D_1 \wedge D_2} A \otimes E \approx E.$$

14. (a) For any derivation D of a ring A into itself, prove **Leibniz's rule**:

$$D^n(xy) = \sum_{i=0}^{n} \binom{n}{i} D^i(x) D^{n-i}(y).$$

(b) Suppose A has characteristic p. Show that D^p is a derivation.

15. Let A/R be an algebra, and let E be an A-module with a connection ∇. Assume that R has characteristic p. Define

$$\psi : \mathrm{Der}(A/R) \to \mathrm{End}_R(E)$$

by

$$\psi(D) = (\nabla(D))^p - \nabla(D^p).$$

Prove that $\psi(D)$ is A-linear. [*Hint*: Use Leibniz's formula and the definition of a connection.] Thus the image of ψ is actually in $\mathrm{End}_A(E)$.

Some Clifford exercises

16. Let $C_g(E)$ be the Clifford algebra as defined in §4. Define $F_i(C_g) = (k + E)^i$, viewing E as embedded in C_g. Define the similar object $F_i(\bigwedge E)$ in the alternating algebra. Then $F_{i+1} \supset F_i$ in both cases, and we define the i-th graded module $\mathrm{gr}_i = F_i/F_{i-1}$. Show that there is a natural (functorial) isomorphism

$$\mathrm{gr}_i(C_g(E)) \xrightarrow{\approx} \mathrm{gr}_i(\bigwedge E).$$

17. Suppose that $k = \mathbf{R}$, so E is a real vector space, which we now assume of even dimension $2m$. We also assume that g is non-degenerate. We omit the index g since the symmetric form is now fixed, and we write C^+, C^- for the spaces of degree 0 and 1 respectively in the $\mathbf{Z}/2\mathbf{Z}$-grading. For elements x, y in C^+ or C^-, define their **supercommutator** to be

$$\{x, y\} = xy - (-1)^{(\deg x)(\deg y)} yx.$$

Show that F_{2m-1} is generated by supercommutators.

18. Still assuming g non-degenerate, let J be an automorphism of (E, g) (i.e. $g(Jx, Jy) = g(x, y)$ for all $x, y \in E$) such that $J^2 = -\mathrm{id}$. Let $E_\mathbf{C} = \mathbf{C} \otimes_\mathbf{R} E$ be the extension of scalars from $\mathbf{R}$ to $\mathbf{C}$. Then $E_\mathbf{C}$ has a direct sum decomposition

$$E_\mathbf{C} = E_\mathbf{C}^+ \oplus E_\mathbf{C}^-$$

into the eigenspaces of J, with eigenvalues 1 and -1 respectively. (Proof?) There is a representation of $E_\mathbf{C}$ on $\bigwedge E_\mathbf{C}^+$, i.e. a homomorphism $E_\mathbf{C} \to \mathrm{End}_\mathbf{C}(E_\mathbf{C}^+)$ whereby an element of $E_\mathbf{C}^+$ operates by exterior multiplication, and an element of $E_\mathbf{C}^-$ operates by inner multiplication, defined as follows.

For $x' \in E_\mathbf{C}^-$ there is a unique $\mathbf{C}$-linear map having the effect

$$x'(x_1 \wedge \cdots \wedge x_r) = -2 \sum_{i=1}^{r} (-1)^{i-1} \langle x', x_i \rangle x_1 \wedge \cdots \wedge \hat{x}_i \wedge \cdots \wedge x_r.$$

Prove that under this operation, you get an isomorphism

$$C_g(E)_{\mathbf{C}} \to \mathrm{End}_{\mathbf{C}}(\bigwedge E^+_{\mathbf{C}}).$$

[*Hint*: Count dimensions.]

19. Consider the Clifford algebra over $\mathbf{R}$. The standard notation is C_n if $E = \mathbf{R}^n$ with the negative definite form, and C'_n if $E = \mathbf{R}^n$ with the positive definite form. Thus $\dim C_n = \dim C'_n = 2^n$.
   (a) Show that

$$C_1 \approx \mathbf{C} \qquad C_2 \approx \mathbf{H} \text{ (the division ring of quaternions)}$$
$$C'_1 \approx \mathbf{R} \times \mathbf{R} \qquad C'_2 \approx M_2(\mathbf{R}) \text{ (}2 \times 2\text{ matrices over }\mathbf{R}\text{)}$$

20. Establish isomorphisms:

$$\mathbf{C} \otimes_{\mathbf{R}} \mathbf{C} \approx \mathbf{C} \times \mathbf{C}; \qquad \mathbf{C} \otimes_{\mathbf{R}} \mathbf{H} \approx M_2(\mathbf{C}); \qquad \mathbf{H} \otimes_{\mathbf{R}} \mathbf{H} \approx M_4(\mathbf{R})$$

where $M_d(F) = d \times d$ matrices over F. For the third one, with $\mathbf{H} \otimes \mathbf{H}$, define an isomorphism

$$f : \mathbf{H} \otimes_{\mathbf{R}} \mathbf{H} \to \mathrm{Hom}_{\mathbf{R}}(\mathbf{H}, \mathbf{H}) \approx M_4(\mathbf{R})$$

by $f(x \otimes y)(z) = xz\bar{y}$, where if $y = y_0 + y_1 i + y_2 j + y_3 k$ then

$$\bar{y} = y_0 - y_1 i - y_2 j - y_3 k.$$

21. (a) Establish isomorphisms

$$C_{n+2} \approx C'_n \otimes C_2 \qquad \text{and} \qquad C'_{n+2} \approx C_n \otimes C'_2.$$

   [*Hint*: Let $\{e_1, \ldots, e_{n+2}\}$ be the orthonormalized basis with $e_i^2 = -1$. Then for the first isomorphism map $e_i \mapsto e'_i \otimes e_1 e_2$ for $i = 1, \ldots, n$ and map e_{n+1}, e_{n+2} on $1 \otimes e_1$ and $1 \otimes e_2$ respectively.]
   (b) Prove that $C_{n+8} \approx C_n \otimes M_{16}(\mathbf{R})$ (which is called the **periodicity property**).
   (c) Conclude that C_n is a semi-simple algebra over $\mathbf{R}$ for all n.

   From (c) one can tabulate the simple modules over C_n. See [ABS 64], reproduced in Husemoller [Hu 75], Chapter 11, §6.

Part Four

HOMOLOGICAL ALGEBRA

In the forties and fifties (mostly in the works of Cartan, Eilenberg, MacLane, and Steenrod, see [CaE 57]), it was realized that there was a systematic way of developing certain relations of linear algebra, depending only on fairly general constructions which were mostly arrow-theoretic, and were affectionately called **abstract nonsense** by Steenrod. (For a more recent text, see [Ro 79].) The results formed a body of algebra, some of it involving homological algebra, which had arisen in topology, algebra, partial differential equations, and algebraic geometry. In topology, some of these constructions had been used in part to get homology and cohomology groups of topological spaces as in Eilenberg-Steenrod [ES 52]. In algebra, factor sets and 1-cocycles had arisen in the theory of group extensions, and, for instance, Hilbert's Theorem 90. More recently, homological algebra has entered in the cohomology of groups and the representation theory of groups. See for example Curtis-Reiner [CuR 81], and any book on the cohomology of groups, e.g. [La 96], [Se 64], and [Sh 72]. Note that [La 96] was written to provide background for class field theory in [ArT 68].

From an entirely different direction, Leray developed a theory of sheaves and spectral sequences motivated by partial differential equations. The basic theory of sheaves was treated in Godement's book on the subject [Go 58]. Fundamental insights were also given by Grothendieck in homological algebra [Gro 57], to be applied by Grothendieck in the theory of sheaves over schemes in the fifties and sixties. In Chapter XX, I have included whatever is necessary of homological algebra for Hartshorne's use in [Ha 77]. Both Chapters XX and XXI give an appropriate background for the homological algebra used in Griffiths-Harris [GrH 78], Chapter 5 (especially §3 and §4), and Gunning [Gu 90]. Chapter XX carries out the general theory of derived functors. The exercises and Chapter XXI may be viewed as providing examples and computations in specific concrete instances of more specialized interest.

The commutative algebra of Chapter X and the two chapters on homological algebra in this fourth part also provide an appropriate background for certain topics in algebraic geometry such as Serre's study of intersection theory [Se 65], Grothendieck duality, and Grothendieck's Riemann-Roch theorem in algebraic geometry. See for instance [SGA 6].

Finally I want to draw attention to the use of homological algebra in certain areas of partial differential equations, as in the papers of Atiyah-Bott-Patodi and Atiyah-Singer on complexes of elliptic operators. Readers can trace some of the literature from the bibliography given in [ABP 73].

The choice of material in this part was to a large extent motivated by all the above applications.

For this chapter, considering the number of references and cross-references given, the bibliography for the entire chapter is placed at the end of the chapter.

CHAPTER XX

General Homology Theory

To a large extent the present chapter is arrow-theoretic. There is a substantial body of linear algebra which can be formalized very systematically, and constitutes what Steenrod called abstract nonsense, but which provides a well-oiled machinery applicable to many domains. References will be given along the way.

Most of what we shall do applies to abelian categories, which were mentioned in Chapter III, end of §3. However, in first reading, I recommend that readers disregard any allusions to general abelian categories and assume that we are dealing with an abelian category of modules over a ring, or other specific abelian categories such as complexes of modules over a ring.

§1. COMPLEXES

Let A be a ring. By an **open complex** of A-modules, one means a sequence of modules and homomorphisms $\{(E^i, d^i)\}$,

$$\to E^{i-1} \xrightarrow{d^{i-1}} E^i \xrightarrow{d^i} E^{i+1} \to$$

where i ranges over all integers and d_i maps E^i into E^{i+1}, and such that

$$d^i \circ d^{i-1} = 0$$

for all i.

One frequently considers a finite sequence of homomorphisms, say

$$E^1 \to \cdots \to E^r$$

such that the composite of two successive ones is 0, and one can make this sequence into a complex by inserting 0 at each end:

$$\to 0 \to 0 \to E^1 \to \cdots \to E^r \to 0 \to 0 \to$$

Such a complex is called a **finite** or **bounded** complex.

Remark. Complexes can be indexed with a descending sequence of integers, namely,

$$\to E_{i+1} \xrightarrow{d_{i+1}} E_i \xrightarrow{d_i} E_{i-1} \to$$

When that notation is used systematically, then one uses upper indices for complexes which are indexed with an ascending sequence of integers:

$$\to E^{i-1} \xrightarrow{d^{i-1}} E^i \xrightarrow{d^i} E^{i+1} \to$$

In this book, I shall deal mostly with ascending indices.

As stated in the introduction of this chapter, instead of modules over a ring, we could have taken objects in an arbitrary abelian category.

The homomorphisms d^i are often called **differentials**, because some of the first complexes which arose in practice were in analysis, with differential operators and differential forms. *Cf.* the examples below.

We denote a complex as above by (E, d). If the complex is exact, it is often useful to insert the kernels and cokernels of the differentials in a diagram as follows, letting $M_i = \text{Ker } d^i = \text{Im } d^{i-1}$.

$$\begin{array}{ccccccccccc}
\longrightarrow & E^{i-2} & \longrightarrow & E^{i-1} & \longrightarrow & E^i & \longrightarrow & E^{i+1} & \longrightarrow \\
& \searrow & \nearrow & & \searrow \quad \nearrow & & \searrow & \nearrow & \\
& & M^{i-1} & & M^i & & M^{i+1} & & \\
& \nearrow & \searrow & & \nearrow \quad \searrow & & \nearrow & \searrow & \\
0 & & & 0 & & 0 & & & 0
\end{array}$$

Thus by definition, we obtain a family of short exact sequences

$$0 \to M^i \to E^i \to M^{i+1} \to 0.$$

If the complex is not exact, then of course we have to insert both the image of d^{i-1} and the kernel of d^i. The factor

$$(\text{Ker } d^i)/(\text{Im } d^{i-1})$$

will be studied in the next section. It is called the **homology of the complex**, and measures the deviation from exactness.

Let M be a module. By a **resolution** of M we mean an exact sequence

$$\to E_n \to E_{n-1} \to \cdots \to E_0 \to M \to 0.$$

Thus a resolution is an exact complex whose furthest term on the right before 0 is M. The resolution is indexed as shown. We usually write E_M for the part of complex formed only of the E_i's, thus:

$$E_M \text{ is:} \quad \to E_n \to E_{n-1} \to \cdots \to E_0,$$

stopping at E_0. We then write E for the complex obtained by sticking 0 on the right:

$$E \text{ is:} \quad \to E_n \to E_{n-1} \to \cdots \to E_0 \to 0.$$

If the objects E_i of the resolution are taken in some family, then the resolution is qualified in the same way as the family. For instance, if E_i is free for all $i \geqq 0$ then we say that the **resolution** is a **free resolution**. If E_i is projective for all $i \geqq 0$ then we say that the **resolution** is **projective**. And so forth. The same terminology is applied to the right, with a resolution

$$0 \to M \to E^0 \to E^1 \to \cdots \to E^{n-1} \to E^n \to,$$

also written

$$0 \to M \to E_M.$$

We then write E for the complex

$$0 \to E^0 \to E^1 \to E^2 \to \cdots.$$

See §5 for injective resolutions.

A resolution is said to be **finite** if E_i (or E^i) $= 0$ for all but a finite number of indices i.

Example. Every module admits a free resolution (on the left). This is a simple application of the notion of free module. Indeed, let M be a module, and let $\{x_j\}$ be a family of generators, with j in some indexing set J. For each j let Re_j be a free module over R with a basis consisting of one element e_j. Let

$$F = \bigoplus_{j \in J} Re_j$$

be their direct sum. There is a unique epimorphism

$$F \to M \to 0$$

sending e_j on x_j, Now we let M_1 be the kernel, and again represent M_1 as the quotient of a free module. Inductively, we can construct the desired free resolution.

Example. The Standard Complex. Let S be a set. For $i = 0, 1, 2, \ldots$ let E_i be the free module over $\mathbf{Z}$ generated by $(i + 1)$-tuples $(x_0, \ldots, x_i)$ with $x_0, \ldots, x_i \in S$. Thus such $(i + 1)$-tuples form a basis of E_i over $\mathbf{Z}$. There is a unique homomorphism

$$d_{i+1} : E_{i+1} \to E_i$$

such that

$$d_{i+1}(x_0, \ldots, x_{i+1}) = \sum_{j=0}^{i+1} (-1)^j (x_0, \ldots, \hat{x}_j, \ldots, x_{i+1}),$$

where the symbol $\hat{x}_j$ means that this term is to be omitted. For $i = 0$, we define $d_0 : E_0 \to \mathbf{Z}$ to be the unique homomorphism such that $d_0(x_0) = 1$. The map d_0 is sometimes called the augmentation, and is also denoted by ε. Then we obtain a resolution of $\mathbf{Z}$ by the complex

$$\to E_{i+1} \to E_i \to \cdots \to E_0 \xrightarrow{\varepsilon} \mathbf{Z} \to 0.$$

The formalism of the above maps d_i is pervasive in mathematics. See Exercise 2 for the use of the standard complex in the cohomology theory of groups. For still another example of this same formalism, compare with the Koszul complex in Chapter XXI, §4.

Given a module M, one may form $\mathrm{Hom}(E_i, M)$ for each i, in which case one gets coboundary maps

$$\delta^i : \mathrm{Hom}(E_i, M) \to \mathrm{Hom}(E_{i+1}, M), \quad \delta(f) = f \circ d^{i+1},$$

obtained by composition of mappings. This procedure will be used to obtain derived functors in §6. In Exercises 2 through 6, you will see how this procedure is used to develop the cohomology theory of groups.

Instead of using homomorphisms, one may use a topological version with simplices, and continuous maps, in which case the standard complex gives rise to the singular homology theory of topological spaces. See [GreH 81], Chapter 9.

Examples. Finite free resolutions. In Chapter XXI, you will find other examples of complexes, especially finite free, constructed in various ways with different tools. This subsequent entire chapter may be viewed as providing examples for the current chapter.

Examples with differential forms. In Chapter XIX, §3, we gave the example of the de Rham complex in an algebraic setting. In the theory of differential manifolds, the de Rham complex has differential maps

$$d^i : \Omega^i \to \Omega^{i+1},$$

sending differential forms of degree i to those of degree $i + 1$, and allows for the computation of the homology of the manifold.

A similar situation occurs in complex differential geometry, when the maps d^i are given by the **Dolbeault** $\bar{\partial}$-operators

$$\bar{\partial}^i : \Omega^{p,i} \to \Omega^{p,i+1}$$

operating on forms of type (p, i). Interested readers can look up for instance Gunning's book [Gu 90] mentioned in the introduction to Part IV, Volume I, E. The associated homology of this complex is called the **Dolbeault** or **$\bar{\partial}$-cohomology** of the complex manifold.

Let us return to the general algebraic aspects of complexes and resolutions.

It is an interesting problem to discuss which modules admit finite resoutions, and variations on this theme. Some conditions are discussed later in this chapter and in Chapter XXI. If a resolution

$$0 \to E_n \to E_{n-1} \to \cdots \to E_0 \to M \to 0$$

is such that $E_m = 0$ for $m > n$, then we say that the resolution has **length** $\leqq n$ (sometimes we say it has **length** n by abuse of language).

A **closed complex** of A-modules is a sequence of modules and homomorphisms $\{(E^i, d^i)\}$ where i ranges over the set of integers mod n for some $n \geqq 2$ and otherwise satisfying the same properties as above. Thus a closed complex looks like this:

$$E^1 \to E^2 \to \cdots \to E^n \quad (E^n \to E^1)$$

We call n the **length** of the closed complex.

Without fear of confusion, one can omit the index i on d^i and write just d. We also write (E, d) for the complex $\{(E^i, d^i)\}$, or even more briefly, we write simply E.

Let (E, d) and (E', d') be complexes (both open or both closed). Let r be an integer. A **morphism** or **homomorphism** (of complexes)

$$f : (E', d') \to (E, d)$$

of **degree** r is a sequence

$$f_i : E'^i \to E^{i+r}$$

of homomorphisms such that for all i the following diagram is commutative:

$$\begin{array}{ccc} E'^{(i-1)} & \xrightarrow{f_{i-1}} & E^{i-1+r} \\ {\scriptstyle d'}\downarrow & & \downarrow{\scriptstyle d} \\ E'^i & \xrightarrow[f_i]{} & E^{i+r} \end{array}$$

Just as we write d instead of d^i, we shall also write f instead of f_i. If the complexes are closed, we define a morphism from one into the other only if they have the same length.

It is clear that complexes form a category. In fact they form an abelian category. Indeed, say we deal with complexes indexed by $\mathbf{Z}$ for simplicity, and morphisms of degree 0. Say we have a morphism of complexes $f : C \to C''$ or

putting the indices:

$$
\begin{array}{ccccccc}
\longrightarrow & C_n & \longrightarrow & C_{n-1} & \longrightarrow \\
& \downarrow & & \downarrow & \\
\longrightarrow & C''_n & \longrightarrow & C''_{n-1} & \longrightarrow
\end{array}
$$

We let $C'_n = \mathrm{Ker}(C_n \to C''_n)$. Then the family (C'_n) forms a complex, which we define to be the kernel of f. We let the reader check the details that this and a similar definition for cokernel and finite direct sums make complexes of modules into an abelian category. At this point, readers should refer to Chapter III, §9, where kernels and cokernels are discussed in this context. The snake lemma of that chapter will now become central to the next section.

It will be useful to have another notion to deal with objects indexed by a monoid. Let G be a monoid, which we assume commutative and additive to fit the applications we have in mind here. Let $\{M_i\}_{i \in G}$ be a family of modules indexed by G. The direct sum

$$M = \bigoplus_{i \in G} M_i$$

will be called the **G-graded module associated with the family** $\{M_i\}_{i \in G}$. Let $\{M_i\}_{i \in G}$ and $\{M'_i\}_{i \in G}$ be families indexed by G, and let M, M' be their associated G-graded modules. Let $r \in G$. By a **G-graded morphism** $f: M' \to M$ of degree r we shall mean a homomorphism such that f maps M'_i into M_{i+r} for each $i \in G$ (identifying M_i with the corresponding submodule of the direct sum on the i-th component). Thus f is nothing else than a family of homomorphisms $f_i : M'_i \to M_{i+r}$.

If (E, d) is a complex we may view E as a G-graded module (taking the direct sum of the components of the complex), and we may view d as a G-graded morphism of degree 1, letting G be $\mathbf{Z}$ or $\mathbf{Z}/n\mathbf{Z}$. The most common case we encounter is when $G = \mathbf{Z}$. Then we write the complex as

$$E = \bigoplus E_i, \quad \text{and} \quad d : E \to E$$

maps E into itself. The differential d is defined as d_i on each direct summand E_i, and has degree 1.

Conversely, if G is $\mathbf{Z}$ or $\mathbf{Z}/n\mathbf{Z}$, one may view a G-graded module as a complex, by defining d to be the zero map.

For simplicity, we shall often omit the prefix "G-graded" in front of the word "morphism", when dealing with G-graded morphisms.

§2. HOMOLOGY SEQUENCE

Let (E, d) be a complex. We let

$$Z^i(E) = \operatorname{Ker} d^i$$

and call $Z^i(E)$ the module of ***i*-cycles**. We let

$$B^i(E) = \operatorname{Im} d^{i-1}$$

and call $B^i(E)$ the module of ***i*-boundaries**. We frequently write Z^i and B^i instead of $Z^i(E)$ and $B^i(E)$, respectively. We let

$$H^i(E) = Z^i/B^i = \operatorname{Ker} d^i/\operatorname{Im} d^{i-1},$$

and call $H^i(E)$ the i-th **homology group** of the complex. The graded module associated with the family $\{H^i\}$ will be denoted by $H(E)$, and will be called the **homology** of E. One sometimes writes $H^*(E)$ instead of $H(E)$.

If $f: E' \to E$ is a morphism of complexes, say of degree 0, then we get an **induced canonical homomorphism**

$$H^i(f) : H^i(E') \to H^i(E)$$

on each homology group. Indeed, from the commutative diagram defining a morphism of complexes, one sees at once that f maps $Z^i(E')$ into $Z^i(E)$ and $B^i(E')$ into $B^i(E)$, whence the induced homomorphism $H^i(f)$. Compare with the beginning remarks of Chapter III, §9. One often writes this induced homomorphism as f_{i*} rather than $H_i(f)$, and if $H(E)$ denotes the graded module of homology as above, then we write

$$H(f) = f_* : H(E') \to H(E).$$

We call $H(f)$ the map **induced** by f on homology. If $H^i(f)$ is an isomorphism for all i, then we say that f is a **homology isomorphism**.

Note that if $f: E' \to E$ and $g: E \to E''$ are morphisms of complexes, then it is immediately verified that

$$H(g) \circ H(f) = H(g \circ f) \quad \text{and} \quad H(\mathrm{id}) = \mathrm{id}.$$

Thus H is a functor from the category of complexes to the category of graded modules.

We shall consider short exact sequences of complexes with morphisms of degree 0:

$$0 \to E' \xrightarrow{f} E \xrightarrow{g} E'' \to 0,$$

which written out in full look like this:

$$\begin{array}{ccccccccc} & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & E'^{(i-1)} & \longrightarrow & E^{i-1} & \longrightarrow & E''^{(i-1)} & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & E'^{i} & \xrightarrow{f} & E^{i} & \xrightarrow{g} & E''^{i} & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & E'^{(i+1)} & \xrightarrow{f} & E^{i+1} & \xrightarrow{g} & E''^{(i+1)} & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & E'^{(i+2)} & \longrightarrow & E^{i+2} & \longrightarrow & E''^{(i+2)} & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \end{array}$$

One can define a morphism

$$\delta : H(E'') \to H(E')$$

of degree 1, *in other words, a family of homomorphisms*

$$\delta^i : H''^i \to H'^{(i+1)}$$

by the snake lemma.

Theorem 2.1. *Let*

$$0 \to E' \xrightarrow{f} E \xrightarrow{g} E'' \to 0$$

be an exact sequence of complexes with f, g of degree 0. *Then the sequence*

$$\begin{array}{ccc} H(E') & \xrightarrow{f_*} & H(E) \\ {\scriptstyle\delta}\nwarrow & & \swarrow{\scriptstyle g_*} \\ & H(E'') & \end{array}$$

is exact.

This theorem is merely a special application of the snake lemma.

If one writes out in full the homology sequence in the theorem, then it looks like this:

$$\boxed{\xrightarrow{\delta} H'^i \to H^i \to H''^i \xrightarrow{\delta} H'^{(i+1)} \to H^{i+1} \to H''^{(i+1)} \xrightarrow{\delta}}$$

It is clear that our map δ is functorial (in an obvious sense), and hence that our whole structure (H, δ) is a functor from the category of short exact sequences of complexes into the category of complexes.

§3. EULER CHARACTERISTIC AND THE GROTHENDIECK GROUP

This section may be viewed as a continuation of Chapter III, §8, on Euler-Poincaré maps. Consider complexes of A-modules, for simplicity.

Let E be a complex such that almost all homology groups H^i are equal to 0. Assume that E is an open complex. As in Chapter III, §8, let φ be an Euler-Poincaré mapping on the category of modules (i.e. A-modules). We define the **Euler-Poincaré characteristic** $\chi_\varphi(E)$ (or more briefly the **Euler characteristic**) with respect to φ, to be

$$\chi_\varphi(E) = \sum (-1)^i \varphi(H^i)$$

provided $\varphi(H^i)$ is defined for all H^i, in which case we say that χ_φ is **defined** for the complex E.

If E is a closed complex, we select a definite order $(E^1, \ldots, E^n)$ for the integers mod n and define the Euler characteristic by the formula

$$\chi_\varphi(E) = \sum_{i=1}^{n} (-1)^i \varphi(H^i)$$

provided again all $\varphi(H^i)$ are defined.

For an example, the reader may refer to Exercise 28 of Chapter I.

One may view H as a complex, defining d to be the zero map. In that case, we see that $\chi_\varphi(H)$ is the alternating sum given above. More generally:

Theorem 3.1. *Let F be a complex, which is of even length if it is closed. Assume that $\varphi(F^i)$ is defined for all i, $\varphi(F^i) = 0$ for almost all i, and $H^i(F) = 0$ for almost all i. Then $\chi_\varphi(F)$ is defined, and*

$$\chi_\varphi(F) = \sum_i (-1)^i \varphi(F^i).$$

Proof. Let Z^i and B^i be the groups of i-cycles and i-boundaries in F^i respectively. We have an exact sequence

$$0 \to Z^i \to F^i \to B^{i+1} \to 0.$$

Hence $\chi_\varphi(F)$ is defined, and

$$\varphi(F^i) = \varphi(Z^i) + \varphi(B^{i+1}).$$

Taking the alternating sum, our conclusion follows at once.

A complex whose homology is trivial is called **acyclic**.

Corollary 3.2. *Let F be an acyclic complex, such that $\varphi(F^i)$ is defined for all i, and equal to 0 for almost all i. If F is closed, we assume that F has even length. Then*

$$\chi_\varphi(F) = 0.$$

In many applications, an open complex F is such that $F^i = 0$ for almost all i, and one can then treat this complex as a closed complex by defining an additional map going from a zero on the far right to a zero on the far left. Thus in this case, the study of such an open complex is reduced to the study of a closed complex.

Theorem 3.3. *Let*

$$0 \to E' \to E \to E'' \to 0$$

be an exact sequence of complexes, with morphisms of degree 0. If the complexes are closed, assume that their length is even. Let φ be an Euler-Poincaré mapping on the category of modules. If χ_φ is defined for two of the above three complexes, then it is defined for the third, and we have

$$\chi_\varphi(E) = \chi_\varphi(E') + \chi_\varphi(E'').$$

Proof. We have an exact homology sequence

$$\to H''^{(i-1)} \to H'^i \to H^i \to H''^i \to H'^{(i+1)} \to$$

This homology sequence is nothing but a complex whose homology is trivial. Furthermore, each homology group belonging say to E is between homology groups of E' and E''. Hence if χ_φ is defined for E' and E'' it is defined for E. Similarly for the other two possibilities. If our complexes are closed of even length n, then this homology sequence has even length $3n$. We can therefore apply the corollary of Theorem 3.1 to get what we want.

For certain applications, it is convenient to construct a universal Euler mapping. Let $\mathfrak{A}$ be the set of isomorphism classes of certain modules. If E is a module, let $[E]$ denote its isomorphism class. We require that $\mathfrak{A}$ satisfy the **Euler-Poincaré condition**, i.e. if we have an exact sequence

$$0 \to E' \to E \to E'' \to 0,$$

then $[E]$ is in $\mathfrak{A}$ if and only if $[E']$ and $[E'']$ are in $\mathfrak{A}$. Furthermore, the zero module is in $\mathfrak{A}$.

Theorem 3.4. *Assume that $\mathfrak{A}$ satisfies the Euler-Poincaré condition. Then there is a map*

$$\gamma : \mathfrak{A} \to \mathbf{K}(\mathfrak{A})$$

of $\mathfrak{A}$ into an abelian group $\mathbf{K}(\mathfrak{A})$ having the universal property with respect to Euler-Poincaré maps defined on $\mathfrak{A}$.

To construct this, let $F_{ab}(\mathfrak{A})$ be the free abelian group generated by the set of such $[E]$. Let B be the subgroup generated by all elements of type

$$[E] - [E'] - [E''],$$

where

$$0 \to E' \to E \to E'' \to 0$$

is an exact sequence whose members are in $\mathfrak{A}$. We let $\mathbf{K}(\mathfrak{A})$ be the factor group $F_{ab}(\mathfrak{A})/B$, and let $\gamma : \mathfrak{A} \to \mathbf{K}(\mathfrak{A})$ be the natural map. It is clear that γ has the universal property.

We observe the similarity of construction with the Grothendieck group of a monoid. In fact, the present group is known as the **Euler-Grothendieck group** of $\mathfrak{A}$, with Euler usually left out.

The reader should observe that the above arguments are valid in abelian categories, although we still used the word **module**. Just as with the elementary isomorphism theorems for groups, we have the analogue of the Jordan-Hölder theorem for modules. Of course in the case of modules, we don't have to worry about the normality of submodules.

We now go a little deeper into **K**-theory. Let $\mathfrak{A}$ be an abelian category. In first reading, one may wish to limit attention to an abelian category of modules over a ring. Let $\mathfrak{C}$ be a family of objects in $\mathfrak{A}$. We shall say that $\mathfrak{C}$ is a **K-family** if it satisfies the following conditions.

K 1. $\mathfrak{C}$ is closed under taking finite direct sums, and 0 is in $\mathfrak{C}$.

K 2. Given an object E in $\mathfrak{A}$ there exists an epimorphism

$$L \to E \to 0$$

with L in $\mathfrak{C}$.

K 3. Let E be an object admitting a finite resolution of length n

$$0 \to L_n \to \cdots \to L_0 \to E \to 0$$

with $L_i \in \mathfrak{C}$ for all i. If

$$0 \to N \to F_{n-1} \to \cdots \to F_0 \to E \to 0$$

is a resolution with N in $\mathfrak{A}$ and $F_0, \ldots, F_{n-1}$ in $\mathfrak{C}$, then N is also in $\mathfrak{C}$.

We note that it follows from these axioms that if F is in $\mathcal{C}$ and F' is isomorphic to F, then F' is also in $\mathcal{C}$, as one sees by looking at the resolution

$$0 \to F' \to F \to 0 \to 0$$

and applying **K 3**. Furthermore, given an exact sequence

$$0 \to F' \to F \to F'' \to 0$$

with F and F'' in $\mathcal{C}$, then F' is in $\mathcal{C}$, again by applying **K 3**.

Example. One may take for $\mathfrak{A}$ the category of modules over a commutative ring, and for $\mathcal{C}$ the family of projective modules. Later we shall also consider Noetherian rings, in which case one may take finite modules, and finite projective modules instead. Condition **K 2** will be discussed in §8.

From now on we assume that $\mathcal{C}$ is a **K**-family. For each object E in $\mathfrak{A}$, we let $[E]$ denote its isomorphism class. An object E of $\mathfrak{A}$ will be said to have **finite $\mathcal{C}$-dimension** if it admits a finite resolution with elements of $\mathcal{C}$. We let $\mathfrak{A}(\mathcal{C})$ be the family of objects in $\mathfrak{A}$ which are of finite $\mathcal{C}$-dimension. We may then form the

$$\mathbf{K}(\mathfrak{A}(\mathcal{C})) = \mathbf{Z}[\mathfrak{A}(\mathcal{C})]/R(\mathfrak{A}(\mathcal{C}))$$

where $R(\mathfrak{A}(\mathcal{C}))$ is the group generated by all elements $[E] - [E'] - [E'']$ arising from an exact sequence

$$0 \to E' \to E \to E'' \to 0$$

in $\mathfrak{A}(\mathcal{C})$. Similarly we define

$$\mathbf{K}(\mathcal{C}) = \mathbf{Z}[(\mathcal{C})]/R(\mathcal{C}),$$

where $R(\mathcal{C})$ is the group of relations generated as above, but taking E', E, E'' in $\mathcal{C}$ itself.

There are natural maps

$$\gamma_{\mathfrak{A}(\mathcal{C})} : \mathfrak{A}(\mathcal{C}) \to \mathbf{K}(\mathfrak{A}(\mathcal{C})) \quad \text{and} \quad \gamma_{\mathcal{C}} : \mathcal{C} \to \mathbf{K}(\mathcal{C}),$$

which to each object associate its class in the corresponding Grothendieck group. There is also a natural homomorphism

$$\epsilon : \mathbf{K}(\mathcal{C}) \to \mathbf{K}(\mathfrak{A}(\mathcal{C}))$$

since an exact sequence of objects of $\mathcal{C}$ can also be viewed as an exact sequence of objects of $\mathfrak{A}(\mathcal{C})$.

Theorem 3.5. *Let $M \in \mathfrak{A}(\mathfrak{C})$ and suppose we have two resolutions*

$$L_M \to M \to 0 \quad \text{and} \quad L'_M \to M \to 0,$$

by finite complexes L_M and L'_M in $\mathfrak{C}$. Then

$$\sum (-1)^i \gamma_{\mathfrak{C}}(L_i) = \sum (-1)^i \gamma_{\mathfrak{C}}(L'_i).$$

Proof. Take first the special case when there is an epimorphism $L'_M \to L_M$, with kernel E illustrated on the following commutative and exact diagram.

$$\begin{array}{ccccccccc}
0 & \longrightarrow & E & \longrightarrow & L'_M & \longrightarrow & L_M & \longrightarrow & 0 \\
 & & & & \downarrow & & \downarrow & & \\
 & & & & M & \xrightarrow{\text{id}} & M & & \\
 & & & & \downarrow & & \downarrow & & \\
 & & & & 0 & & 0 & &
\end{array}$$

The kernel is a complex

$$0 \to E_n \to E_{n-1} \to \cdots \to E_0 \to 0$$

which is exact because we have the homology sequence

$$H_p(E) \to H_p(L') \to H_p(L) \to H_{p-1}(E)$$

For $p \geqq 1$ we have $H_p(L) = H_p(L') = 0$ by definition, so $H_p(E) = 0$ for $p \geqq 1$. And for $p = 0$ we consider the exact sequence

$$H_1(L) \to H_0(E) \to H_0(L') \to H_0(L)$$

Now we have $H_1(L) = 0$, and $H_0(L') \to H_0(L)$ corresponds to the identity morphisms on M so is an isomorphism. It follows that $H_0(E) = 0$ also.

By definition of **K**-family, the objects E_p are in $\mathfrak{C}$. Then taking the Euler characteristic in $\mathbf{K}(\mathfrak{C})$ we find

$$\chi(L') - \chi(L) = \chi(E) = 0$$

which proves our assertion in the special case.

The general case follows by showing that given two resolutions of M in $\mathfrak{C}$ we can always find a third one which tops both of them. The pattern of our construction will be given by a lemma.

Lemma 3.6. *Given two epimorphisms $u: M \to N$ and $v: M' \to N$ in $\mathcal{A}$, there exist epimorphisms $F \to M$ and $F \to M'$ with F in $\mathcal{C}$ making the following diagram commutative.*

$$\begin{array}{ccccc} & & F & & \\ & \swarrow & & \searrow & \\ M & & & & M' \\ & \searrow & & \swarrow & \\ & & N & & \end{array}$$

Proof. Let $E = M \times_N M'$, that is E is the kernel of the morphism

$$M \times M' \to N$$

given by $(x, y) \mapsto ux - vy$. (Elements are not really used here, and we could write formally $u - v$ instead.) There is some F in $\mathcal{C}$ and an epimorphism $F \to E \to 0$. The composition of this epimorphism with the natural projections of E on each factor gives us what we want.

We construct a complex L''_M giving a resolution of M with a commutative and exact diagram:

$$\begin{array}{ccccc} 0 & & & & \\ \uparrow & & & & \\ L_M & \longrightarrow & M & \longrightarrow & 0 \\ \uparrow & & \uparrow \text{id} & & \\ L''_M & \longrightarrow & M & \longrightarrow & 0 \\ \downarrow & & \downarrow \text{id} & & \\ L'_M & \longrightarrow & M & \longrightarrow & 0 \\ \downarrow & & & & \\ 0 & & & & \end{array}$$

The construction is done inductively, so we put indices:

$$\begin{array}{ccccc} L_i & \longrightarrow & L_{i-1} & \longrightarrow & \\ \uparrow & & \uparrow & & \\ L''_i & \longrightarrow & L''_{i-1} & \longrightarrow & \\ \downarrow & & \downarrow & & \\ L'_i & \longrightarrow & L'_{i-1} & \longrightarrow & \end{array}$$

Suppose that we have constructed up to L''_{i-1} with the desired epimorphisms on L_{i-1} and L'_{i-1}. We want to construct L''_i. Let $B_i = \text{Ker}(L_{i-1} \to L_{i-2})$ and similarly for B'_i and B''_i. We obtain the commutative diagram:

$$\begin{array}{ccccccc}
L_i & \longrightarrow & B_i & \longrightarrow & L_{i-1} & \longrightarrow & L_{i-2} \\
 & & \uparrow & & \uparrow & & \uparrow \\
 & & B''_i & \longrightarrow & L''_{i-1} & \longrightarrow & L''_{i-2} \\
 & & \downarrow & & \downarrow & & \downarrow \\
L'_i & \longrightarrow & B'_i & \longrightarrow & L'_{i-1} & \longrightarrow & L'_{i-2}
\end{array}$$

If $B''_i \to B_i$ or $B''_i \to B'_i$ are not epimorphisms, then we replace L''_{i-1} by

$$L''_{i-1} \oplus L_i \oplus L'_i.$$

We let the boundary map to L''_{i-2} be 0 on the new summands, and similarly define the maps to L_{i-1} and L'_{i-1} to be 0 on L'_i and L_{i-1} respectively.

Without loss of generality we may now assume that

$$B''_i \to B_i \quad \text{and} \quad B''_i \to B'_i$$

are epimorphisms. We then use the construction of the preceding lemma. We let

$$E_i = L_i \bigoplus_{B_i} B''_i \quad \text{and} \quad E'_i = B''_i \bigoplus_{B'_i} L'_i.$$

Then both E_i and E'_i have natural epimorphisms on B''_i. Then we let

$$N_i = E_i \bigoplus_{B''_i} E'_i$$

and we find an object L''_i in $\mathcal{C}$ with an epimorphism $L''_i \to N_i$. This gives us the inductive construction of L'' up to the very end. To stop the process, we use **K 3** and take the kernel of the last constructed L''_i to conclude the proof.

Theorem 3.7. *The natural map*

$$\epsilon : \mathbf{K}(\mathcal{C}) \to \mathbf{K}(\mathcal{A}(\mathcal{C}))$$

is an isomorphism.

Proof. The map is surjective because given a resolution

$$0 \to F_n \to \cdots \to F_0 \to M \to 0$$

with $F_i \in \mathcal{C}$ for all i, the element

$$\sum (-1)^i \gamma_{\mathcal{C}}(F_i)$$

maps on $\gamma_{\mathfrak{a}(\mathfrak{C})}(M)$ under ϵ. Conversely, Theorem 3.5 shows that the association

$$M \mapsto \sum (-1)^i \gamma_{\mathfrak{C}}(F_i)$$

is a well-defined mapping. Since for any $L \in \mathfrak{C}$ we have a short exact sequence $0 \to L \to L \to 0$, it follows that this mapping following ϵ is the identity on $\mathbf{K}(\mathfrak{C})$, so ϵ is a monomorphism. Hence ϵ is an isomorphism, as was to be shown.

It may be helpful to the reader actually to see the next lemma which makes the additivity of the inverse more explicit.

Lemma 3.8. *Given an exact sequence in* $\mathfrak{a}(\mathfrak{C})$

$$0 \to M' \to M \to M'' \to 0$$

there exists a commutative and exact diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & L_{M'} & \longrightarrow & L_M & \longrightarrow & L_{M''} & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & M' & \longrightarrow & M & \longrightarrow & M'' & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ & & 0 & & 0 & & 0 & & \end{array}$$

with finite resolutions $L_{M'}$, L_M, $L_{M''}$ in $\mathfrak{C}$.

Proof. We first show that we can find L', L, L'' in $\mathfrak{C}$ to fit an exact and commutative diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & L' & \longrightarrow & L & \longrightarrow & L'' & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & M' & \longrightarrow & M & \longrightarrow & M'' & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ & & 0 & & 0 & & 0 & & \end{array}$$

We first select an epimorphism $L'' \to M''$ with L'' in $\mathfrak{C}$. By Lemma 3.6 there exists $L_1 \in \mathfrak{C}$ and epimorphisms $L_1 \to M$, $L_1 \to L''$ making the diagram commutative. Then let $L_2 \to M'$ be an epimorphism with $L_2 \in \mathfrak{C}$, and finally define $L = L_1 \oplus L_2$. Then we get morphisms $L \to M$ and $L \to L''$ in the obvious way. Let L' be the kernel of $L \to L''$. Then $L_2 \subset L'$ so we get an epimorphism $L' \to M'$.

This now allows us to construct resolutions inductively until we hit the n-th step, where n is some integer such that M, M'' admit resolutions of length n in $\mathcal{C}$. The last horizontal exact sequence that we obtain is

$$0 \to L'_n \to L_n \to L''_n \to 0$$

and L''_n can be chosen to be the kernel of $L''_{n-1} \to L''_{n-2}$. By **K 3** we know that L''_n lies in $\mathcal{C}$, and the sequence

$$0 \to L''_n \to L''_{n-1}$$

is exact. This implies that in the next inductive step, we can take $L''_{n+1} = 0$. Then

$$0 \to L'_{n+1} \to L_{n+1} \to 0 \to 0$$

is exact, and at the next step we just take the kernels of the vertical arrows to complete the desired finite resolutions in $\mathcal{C}$. This concludes the proof of the lemma.

Remark. The argument in the proof of Lemma 3.8 in fact shows:

If

$$0 \to M' \to M \to M'' \to 0$$

is an exact sequence in $\mathfrak{A}$, *and if* M, M'' *have finite* $\mathcal{C}$*-dimension, then so does* M'.

In the category of modules, one has a more precise statement:

Theorem 3.9. *Let* $\mathfrak{A}$ *be the category of modules over a ring. Let* $\mathcal{P}$ *be the family of projective modules. Given an exact sequence of modules*

$$0 \to E' \to E \to E'' \to 0$$

if any two of E', E, E'' *admit finite resolutions in* $\mathcal{P}$ *then the third does also.*

Proofs in a more subtle case will be given in Chapter XXI, Theorem 2.7.

Next we shall use the tensor product to investigate a ring structure on the Grothendieck group. We suppose for simplicity that we deal with an abelian category of modules over a commutative ring, denoted by $\mathfrak{A}$, together with a **K**-family $\mathcal{C}$ as above, but we now assume that $\mathfrak{A}$ is closed under the tensor product. The only properties we shall actually use for the next results are the following ones, denoted by **TG** (for "tensor" and "Grothendieck" respectively):

TG 1. There is a bifunctorial isomorphism giving commutativity

$$M \otimes N \approx N \otimes M$$

for all M, N in $\mathfrak{A}$; and similarly for distributivity over direct sums, and associativity.

TG 2. For all L in $\mathcal{C}$ the functor $M \mapsto L \otimes M$ is exact.

TG 3. If L, L' are in $\mathcal{C}$ then $L \otimes L'$ is in $\mathcal{C}$.

Then we may give $\mathbf{K}(\mathcal{C})$ the structure of an algebra by defining

$$\mathrm{cl}_{\mathcal{C}}(L)\,\mathrm{cl}_{\mathcal{C}}(L') = \mathrm{cl}_{\mathcal{C}}(L \otimes L').$$

Condition **TG 1** implies that this algebra is commutative, and we call it the **Grothendieck algebra**. In practice, there is a unit element, but if we want one in the present axiomatization, we have to make it an explicit assumption:

TG 4. There is an object R in $\mathcal{C}$ such that $R \otimes M \approx M$ for all M in $\mathfrak{A}$.

Then $\mathrm{cl}_{\mathcal{C}}(R)$ is the unit element.

Similarly, condition **TG 2** shows that we can define a module structure on $\mathbf{K}(\mathfrak{A})$ over $\mathbf{K}(\mathcal{C})$ by the same formula

$$\mathrm{cl}_{\mathcal{C}}(L)\,\mathrm{cl}_{\mathfrak{a}}(M) = \mathrm{cl}_{\mathfrak{a}}(L \otimes M),$$

and similarly $\mathbf{K}(\mathfrak{A}(\mathcal{C}))$ is a module over $\mathbf{K}(\mathcal{C})$, where we recall that $\mathfrak{A}(\mathcal{C})$ is the family of objects in $\mathfrak{A}$ which admit finite resolutions by objects in $\mathcal{C}$.

Since we know from Theorem 3.7 that $\mathbf{K}(\mathcal{C}) \approx \mathbf{K}(\mathfrak{A}(\mathcal{C}))$, we also have a ring structure on $\mathbf{K}(\mathfrak{A}(\mathcal{C}))$ via this isomorphism. We then can make the product more explicit as follows.

Proposition 3.10. *Let $M \in \mathfrak{A}(\mathcal{C})$ and let $N \in \mathfrak{A}$. Let*

$$0 \to L_n \to \cdots \to L_0 \to M \to 0$$

be a finite resolution of M by objects in $\mathcal{C}$. Then

$$\begin{aligned}\mathrm{cl}_{\mathcal{C}}(M)\,\mathrm{cl}_{\mathfrak{a}}(N) &= \sum (-1)^i\,\mathrm{cl}_{\mathfrak{a}}(L_i \otimes N).\\ &= \sum (-1)^i\,\mathrm{cl}_{\mathfrak{a}}(H_i(K))\end{aligned}$$

where K is the complex

$$0 \to L_n \otimes N \to \cdots \to L_0 \otimes N \to M \otimes N \to 0$$

and $H_i(K)$ is the i-th homology of this complex.

Proof. The formulas are immediate consequences of the definitions, and of Theorem 3.1.

Example. Let $\mathfrak{A}$ be the abelian category of modules over a commutative ring. Let $\mathcal{C}$ be the family of projective modules. From §6 on derived functors the reader will know that the homology of the complex K in Proposition 3.10 is just $\mathrm{Tor}(M, N)$. Therefore the formula in that proposition can also be written

$$\mathrm{cl}_{\mathcal{C}}(M)\,\mathrm{cl}_{\mathfrak{a}}(N) = \sum (-1)^i\,\mathrm{cl}_{\mathfrak{a}}(\mathrm{Tor}_i(M, N)).$$

Example. Let k be a field. Let G be a group. By a **(G, k)-module**, we shall mean a pair (E, ρ), consisting of a k-space E and a homomorphism

$$\rho: G \to \text{Aut}_k(E).$$

Such a homomorphism is also called a **representation** of G in E. By abuse of language, we also say that the k-space E is a G-module. The group G operates on E, and we write σx instead of $\rho(\sigma)x$. The field k will be kept fixed in what follows.

Let $\text{Mod}_k(G)$ denote the category whose objects are (G, k)-modules. A morphism in $\text{Mod}_k(G)$ is what we call a **G-homomorphism**, that is a k-linear map $f: E \to F$ such that $f(\sigma x) = \sigma f(x)$ for all $\sigma \in G$. The group of morphisms in $\text{Mod}_k(G)$ is denoted by Hom_G.

If E is a G-module, and $\sigma \in G$, then we have by definition a k-automorphism $\sigma : E \to E$. Since T^r is a functor, we have an induced automorphism

$$T^r(\sigma) : T^r(E) \to T^r(E)$$

for each r, and thus $T^r(E)$ is also a G-module. Taking the direct sum, we see that $T(E)$ is a G-module, and hence that T is a functor from the category of G-modules to the category of graded G-modules. Similarly for $\bigwedge^r$, S^r, and $\bigwedge$, S.

It is clear that the kernel of a G-homomorphism is a G-submodule, and that the factor module of a G-module by a G-submodule is again a G-module so the category of G-modules is an abelian category.

We can now apply the general considerations on the Grothendieck group which we write

$$\mathbf{K}(G) = \mathbf{K}(\text{Mod}_k(G))$$

for simplicity in the present case. We have the canonical map

$$\text{cl}: \text{Mod}_k(G) \to \mathbf{K}(G).$$

which to each G-module associates its class in $\mathbf{K}(G)$.

If E, F are G-modules, then their tensor product over k, $E \otimes F$, is also a G-module. Here again, the operation of G on $E \otimes F$ is given functorially. If $\sigma \in G$, there exists a unique k-linear map $E \otimes F \to E \otimes F$ such that for $x \in E$, $y \in F$ we have $x \otimes y \mapsto (\sigma x) \otimes (\sigma y)$. The tensor product induces a law of composition on $\text{Mod}_k(G)$ because the tensor products of G-isomorphic modules are G-isomorphic.

Furthermore all the conditions **TG 1** through **TG 4** are satisfied. Since k is a field, we find also that tensoring an exact sequence of G-modules over k with any G-module over k preserves the exactness, so **TG 2** is satisfied for all (G, k)-modules. Thus the Grothendieck group $\mathbf{K}(G)$ is in fact the Grothendieck ring, or the Grothendieck algebra over k.

By Proposition 2.1 and Theorem 2.3 of Chapter XVIII, we also see:

The Grothendieck ring of a finite group G consisting of isomorphism classes of finite dimensional (G, k)-spaces over a field k of characteristic 0 is naturally isomorphic to the character ring $X_{\mathbf{Z}}(G)$.

We can axiomatize this a little more. We consider an abelian category of modules over a commutative ring R, which we denote by $\mathfrak{A}$ for simplicity. For two modules M, N in $\mathfrak{A}$ we let $\mathrm{Mor}(M, N)$ as usual be the morphisms in $\mathfrak{A}$, but $\mathrm{Mor}(M, N)$ is an abelian subgroup of $\mathrm{Hom}_R(M, N)$. For example, we could take $\mathfrak{A}$ to be the category of (G, k)-modules as in the example we have just discussed, in which case $\mathrm{Mor}(M, N) = \mathrm{Hom}_G(M, N)$.

We let $\mathcal{C}$ be the family of finite free modules in $\mathfrak{A}$. *We assume that $\mathcal{C}$ satisfies* **TG 1, TG 2, TG 3, TG 4,** *and also that $\mathcal{C}$ is closed under taking alternating products, tensor products and symmetric products.* We let $\mathbf{K} = \mathbf{K}(\mathcal{C})$. As we have seen, $\mathbf{K}$ is itself a commutative ring. We abbreviate $\mathrm{cl}_{\mathcal{C}} = \mathrm{cl}$.

We shall define non-linear maps

$$\lambda^i : \mathbf{K} \to \mathbf{K}$$

using the alternating product. If E is finite free, we let

$$\lambda^i(E) = \mathrm{cl}(\textstyle\bigwedge^i E).$$

Proposition 1.1 of Chapter XIX can now be formulated for the $\mathbf{K}$-ring as follows.

Proposition 3.11. *Let*

$$0 \to E' \to E \to E'' \to 0$$

be an exact sequence of finite free modules in $\mathfrak{A}$. Then for every integer $n \geqq 0$ we have

$$\lambda^n(E) = \sum_{i=0}^{n} \lambda^i(E')\lambda^{n-i}(E'').$$

As a result of the proposition, we can define a map

$$\lambda_t : \mathbf{K} \to 1 + t\mathbf{K}[[t]]$$

of $\mathbf{K}$ into the multiplicative group of formal power series with coefficients in $\mathbf{K}$, and with constant term 1, by letting

$$\lambda_t(x) = \sum_{i=0}^{\infty} \lambda^i(x)t^i.$$

Proposition 1.4 of Chapter XIX can be formulated by saying that:

The map

$$\lambda_t : \mathbf{K} \to 1 + t\mathbf{K}[[t]]$$

is a homomorphism.

We note that if L is free of rank 1, *then*

$$\lambda^0(L) = \text{ground ring};$$

$$\lambda^1(L) = \text{cl}(L);$$

$$\lambda^i(L) = 0 \quad \text{for} \quad i > 1.$$

This can be summarized by writing

$$\lambda_t(L) = 1 + \text{cl}(L)t.$$

Next we can do a similar construction with the symmetric product instead of the alternating product. If E is a finite free module in $\mathcal{C}$ we let as usual:

$S(E) =$ symmetric algebra of E;

$S^i(E) =$ homogeneous component of degree i in $S(E)$.

We define

$$\sigma^i(E) = \text{cl}(S^i(E))$$

and the corresponding power series

$$\sigma_t(E) = \sum \sigma^i(E)t^i.$$

Theorem 3.12. *Let E be a finite free module in* $\mathfrak{A}$, *of rank r. Then for all integers* $n \geqq 1$ *we have*

$$\sum_{i=0}^{r} (-1)^i \lambda^i(E)\sigma^{n-i}(E) = 0,$$

where by definition $\sigma^j(E) = 0$ *for* $j < 0$. *Furthermore*

$$\sigma_t(E)\lambda_{-t}(E) = 1,$$

so the power series $\sigma_t(E)$ *and* $\lambda_{-t}(E)$ *are inverse to each other.*

Proof. The first formula depends on the analogue for the symmetric product and the alternating product of the formula given in Proposition 1.1 of Chapter

XIX. It could be proved directly now, but the reader will find a proof as a special case of the theory of Koszul complexes in Chapter XXI, Corollary 4.14. The power series relation is essentially a reformulation of the first formula.

From the above formalism, it is possible to define other maps besides λ^i and σ^i.

Example. Assume that the group G is trivial, and just write $\mathbf{K}$ for the Grothendieck ring instead of $\mathbf{K}(1)$. For $x \in \mathbf{K}$ define

$$\psi_{-t}(x) = -t \frac{d}{dt} \log \lambda_t(x) = -t\, \lambda_t'(x)/\lambda_t(x).$$

Show that ψ_{-t} is an additive and multiplicative homomorphism. Show that

$$\psi_t(E) = 1 + \mathrm{cl}(E)t + \mathrm{cl}(E)^2t^2 + \cdots .$$

This kind of construction with the logarithmic derivative leads to the **Adams operations** ψ^i in topology and algebraic geometry. See Exercise 22 of Chapter XVIII.

Remark. If it happens in Theorem 3.12 that E admits a decomposition into 1-dimensional free modules in the $\mathbf{K}$-group, then the proof trivializes by using the fact that $\lambda_t(L) = 1 + \mathrm{cl}(L)t$ if L is 1-dimensional. But in the example of (G, k)-spaces when k is a field, this is in general not possible, and it is also not possible in other examples arising naturally in topology and algebraic geometry. However, by "changing the base," one can sometimes achieve this simpler situation, but Theorem 3.12 is then used in establishing the basic properties. *Cf.* Grothendieck [SGA 6], mentioned in the introduction to Part IV, and other works mentioned in the bibliography at the end, namely [Ma 69], [At 61], [At 67], [Ba 68], [Bo 62]. The lectures by Atiyah and Bott emphasize the topological aspects as distinguished from the algebraic-geometric aspects. Grothendieck [Gr 68] actually shows how the formalism of Chern classes from algebraic geometry and topology also enters the theory of representations of linear groups. See also the exposition in [FuL 85], especially the formalism of Chapter I, §6. For special emphasis on applications to representation theory, see Bröcker-tom Dieck [BtD 85], especially Chapter II, §7, concerning compact Lie groups.

§4. INJECTIVE MODULES

In Chapter III, §4, we defined projective modules, which have a natural relation to free modules. By reversing the arrows, we can define a module Q to be **injective** if it satisfies any one of the following conditions which are equivalent:

I 1. Given any module M and a submodule M', and a homomorphism $f: M' \to Q$, there exists an extension of this homomorphism to M,

that is there exists $h: M \to Q$ making the following diagram commutative:

$$\begin{array}{ccccc} 0 & \longrightarrow & M' & \longrightarrow & M \\ & & {\scriptstyle f}\downarrow & \swarrow {\scriptstyle h} & \\ & & Q & & \end{array}$$

I 2. The functor $M \mapsto \mathrm{Hom}_A(M, Q)$ is exact.

I 3. Every exact sequence $0 \to Q \to M \to M'' \to 0$ splits.

We prove the equivalence. General considerations on homomorphisms as in Proposition 2.1, show that exactness of the homed sequence may fail only at one point, namely given

$$0 \to M' \to M \to M'' \to 0,$$

the question is whether

$$\mathrm{Hom}_A(M, Q) \to \mathrm{Hom}_A(M', Q) \to 0$$

is exact. But this is precisely the hypothesis as formulated in **I 1**, so **I 1** implies **I 2** is essentially a matter of linguistic reformulation, and in fact **I 1** is equivalent to **I 2**.

Assume **I 2** or **I 1**, which we know are equivalent. To get **I 3** is immediate, by applying **I 1** to the diagram:

$$\begin{array}{ccccc} 0 & \longrightarrow & Q & \longrightarrow & M \\ & & {\scriptstyle \mathrm{id}}\downarrow & \swarrow {\scriptstyle h} & \\ & & Q & & \end{array}$$

To prove the converse, we need the notion of push-out (cf. Exercise 52 of Chapter I). Given an exact diagram

$$\begin{array}{ccccc} 0 & \longrightarrow & M' & \longrightarrow & M \\ & & \downarrow & & \\ & & Q & & \end{array}$$

we form the push-out:

$$\begin{array}{ccc} M' & \longrightarrow & M \\ \downarrow & & \downarrow \\ Q & \longrightarrow & N = Q \oplus_{M'} M. \end{array}$$

Since $M' \to M$ is a monomorphism, it is immediately verified from the construction of the push-out that $Q \to N$ is also a monomorphism. By **I 3**, the sequence

$$0 \to Q \to N$$

splits, and we can now compose the splitting map $N \to Q$ with the push-out map $M \to N$ to get the desired $h: M \to Q$, thus proving **I 1**.

We saw easily that every module is a homomorphic image of a free module. There is no equally direct construction for the dual fact:

Theorem 4.1. *Every module is a submodule of an injective module.*

The proof will be given by dualizing the situation, with some lemmas. We first look at the situation in the category of abelian groups. If M is an abelian group, let its dual group be $M^\wedge = \text{Hom}(M, \mathbf{Q}/\mathbf{Z})$. If F is a free abelian group, it is reasonable to expect, and in fact it is easily proved that its dual $F^\wedge$ is an injective module, since injectivity is the dual notion of projectivity. Furthermore, M has a natural map into the double dual $M^{\wedge\wedge}$, which is shown to be a monomorphism. Now represent $M^\wedge$ as a quotient of a free abelian group,

$$F \to M^\wedge \to 0.$$

Dualizing this sequence yields a monomorphism

$$0 \to M^{\wedge\wedge} \to F^\wedge,$$

and since M is embedded naturally as a subgroup of $M^{\wedge\wedge}$, we get the desired embedding of M as a subgroup of $F^\wedge$.

This proof also works in general, but there are details to be filled in. First we have to prove that the dual of a free module is injective, and second we have to be careful when passing from the category of abelian groups to the category of modules over an arbitrary ring. We now carry out the details.

We say that an abelian group T is **divisible** if for every integer m, the homomorphism

$$m_T : x \mapsto mx$$

is surjective.

Lemma 4.2. *If T is divisible, then T is injective in the category of abelian groups.*

Proof. Let $M' \subset M$ be a subgroup of an abelian group, and let $f: M' \to T$ be a homomorphism. Let $x \in M$. We want first to extend f to the module (M', x) generated by M' and x. If x is free over M', then we select any value $t \in T$, and it is immediately verified that f extends to (M', x) by giving the value $f(x) = t$. Suppose that x is torsion with respect to M', that is there is a positive integer m such that $mx \in M'$. Let d be the period of x mod M', so

$dx \in M'$, and d is the least positive integer such that $dx \in M'$. By hypothesis, there exists an element $u \in T$ such that $du = f(dx)$. For any integer n, and $z \in M'$ define

$$f(z + nx) = f(z) + nu.$$

By the definition of d, and the fact that $\mathbf{Z}$ is principal, one sees that this value for f is independent of the representation of an element of (M', x) in the form $z + nx$, and then it follows at once that this extended definition of f is a homomorphism. Thus we have extended f to (M', x).

The rest of the proof is merely an application of Zorn's lemma. We consider pairs (N, g) consisting of submodules of M containing M', and an extension g of f to N. We say that $(N, g) \leqq (N_1, g_1)$ if $N \subset N_1$ and the restriction of g_1 to N is g. Then such pairs are inductively ordered. Let (N, g) be a maximal element. If $N \neq M$ then there is some $x \in M$, $x \notin N$ and we can apply the first part of the proof to extend the homomorphism to (N, x), which contradicts the maximality, and concludes the proof of the lemma.

Example. The abelian groups $\mathbf{Q}/\mathbf{Z}$ and $\mathbf{R}/\mathbf{Z}$ are divisible, and hence are injective in the category of abelian groups.

We can prove Theorem 4.1 in the category of abelian groups following the pattern described above. If F is a free abelian group, then the dual $F^\wedge$ is a direct product of groups isomorphic to $\mathbf{Q}/\mathbf{Z}$, and is therefore injective in the category of abelian groups by Lemma 4.2. This concludes the proof.

Next we must make the necessary remarks to extend the system to modules. Let A be a ring and let T be an abelian group. We make $\text{Hom}_{\mathbf{Z}}(A, T)$ into an A-module as follows. Let $f: A \to T$ be an abelian group homomorphism. For $a \in A$ we define the operation

$$(af)(b) = f(ba).$$

The rules for an operation are then immediately verified. Then for any A-module X we have a natural isomorphism of abelian groups:

$$\boxed{\text{Hom}_{\mathbf{Z}}(X, T) \xrightarrow{\approx} \text{Hom}_A(X, \text{Hom}_{\mathbf{Z}}(A, T)).}$$

Indeed, let $\psi : X \to T$ be a $\mathbf{Z}$-homomorphism. We associate with ψ the homomorphism

$$f: X \to \text{Hom}_{\mathbf{Z}}(A, T)$$

such that

$$f(x)(a) = \psi(ax).$$

The definition of the A-module structure on $\mathrm{Hom}_{\mathbf{Z}}(A, T)$ shows that f is an A-homomorphism, so we get an arrow from $\mathrm{Hom}_{\mathbf{Z}}(X, T)$ to

$$\mathrm{Hom}_A(X, \mathrm{Hom}_{\mathbf{Z}}(A, T)).$$

Conversely, let $f: X \to \mathrm{Hom}_{\mathbf{Z}}(A, T)$ be an A-homomorphism. We define the corresponding ψ by

$$\psi(x) = f(x)(1).$$

It is then immediately verified that these maps are inverse to each other.

We shall apply this when T is any divisible group, although we think of T as being $\mathbf{Q}/\mathbf{Z}$, and we think of the homomorphisms into T as representing the dual group according to the pattern described previously.

Lemma 4.3. *If T is a divisible abelian group, then* $\mathrm{Hom}_{\mathbf{Z}}(A, T)$ *is injective in the category of A-modules.*

Proof. It suffices to prove that if $0 \to X \to Y$ is exact in the category of A-modules, then the dual sequence obtained by taking A-homomorphisms into $\mathrm{Hom}_{\mathbf{Z}}(A, T)$ is exact, that is the top map in the following diagram is surjective.

$$\begin{array}{ccccc}
\mathrm{Hom}_A(Y, \mathrm{Hom}_{\mathbf{Z}}(A, T)) & \longrightarrow & \mathrm{Hom}_A(X, \mathrm{Hom}_{\mathbf{Z}}(A, T)) & \xrightarrow{?} & 0 \\
\uparrow \approx & & \uparrow \approx & & \\
\mathrm{Hom}_{\mathbf{Z}}(Y, T) & \longrightarrow & \mathrm{Hom}_{\mathbf{Z}}(X, T) & \longrightarrow & 0
\end{array}$$

But we have the isomorphisms described before the lemma, given by the vertical arrows of the diagram, which is commutative. The bottom map is surjective because T is an injective module in the category of abelian groups. Therefore the top map is surjective, thus proving the lemma.

Now we prove Theorem 4.1 for A-modules. Let M be an A-module. We can embed M in a divisible abelian group T,

$$0 \to M \xrightarrow{f} T.$$

Then we get an A-homomorphism

$$M \to \mathrm{Hom}_{\mathbf{Z}}(A, T)$$

by $x \mapsto f_x$, where $f_x(a) = f(ax)$. One verifies at once that $x \mapsto f_x$ gives an embedding of M in $\mathrm{Hom}_{\mathbf{Z}}(A, T)$, which is an injective module by Lemma 4.3. This concludes the proof of Theorem 4.1.

§5. HOMOTOPIES OF MORPHISMS OF COMPLEXES

The purpose of this section is to describe a condition under which homomorphisms of complexes induce the same map on the homology and to show that this condition is satisfied in an important case, from which we derive applications in the next section.

The arguments are applicable to any abelian category. The reader may prefer to think of modules, but we use a language which applies to both, and is no more complicated than if we insisted on dealing only with modules.

Let $E = \{(E^n, d^n)\}$ and $E' = \{(E'^n, d'^n)\}$ be two complexes. Let

$$f, g : E \to E'$$

be two morphisms of complexes (of degree 0). We say that f is **homotopic to** g if there exists a sequence of homomorphisms

$$h_n : E^n \to E'^{(n-1)}$$

such that

$$f_n - g_n = d'^{(n-1)} h_n + h_{n+1} d^n.$$

Lemma 5.1. *If f, g are homotopic, then f, g induce the same homomorphism on the homology $H(E)$, that is*

$$H(f_n) = H(g_n) : H^n(E) \to H^n(E').$$

Proof. The lemma is immediate, because $f_n - g_n$ vanishes on the cycles, which are the kernel of d^n, and the homotopy condition shows that the image of $f_n - g_n$ is contained in the boundaries, that is, in the image of $d'^{(n-1)}$.

Remark. The terminology of homotopy is used because the notion and formalism first arose in the context of topology. Cf. [ES 52] and [GreH 81].

We apply Lemma 5.1 to injective objects. Note that as usual the definition of an injective module applies without change to define an injective object in any abelian category. Instead of a submodule in **I 1**, we use a subobject, or equivalently a monomorphism. The proofs of the equivalence of the three conditions defining an injective module depended only on arrow-theoretic juggling, and apply in the general case of abelian categories.

We say that an abelian category has **enough injectives** if given any object M there exists a monomorphism

$$0 \to M \to I$$

into an injective object. We proved in §4 that the category of modules over a ring has enough injectives. *We now assume that the abelian category we work with has enough injectives.*

By an **injective resolution** of an object M one means an exact sequence

$$0 \to M \to I^0 \to I^1 \to I^2 \to$$

such that each I_n ($n \geqq 0$) is injective. Given M, such a resolution exists. Indeed, the monomorphism

$$0 \to M \to I^0$$

exists by hypothesis. Let M^0 be its image. Again by assumption, there exists a monomorphism

$$0 \to I^0/M^0 \to I^1,$$

and the corresponding homomorphism $I^0 \to I^1$ has kernel M^0. So we have constructed the first step of the resolution, and the next steps proceed in the same fashion.

An injective resolution is of course not unique, but it has some uniqueness which we now formulate.

Lemma 5.2. *Consider two complexes:*

$$\begin{array}{ccccccccccc} 0 & \longrightarrow & M & \longrightarrow & E^0 & \longrightarrow & E^1 & \longrightarrow & E^2 & \longrightarrow & \cdots \\ & & \Big\downarrow{\scriptstyle \varphi} & & & & & & & & \\ 0 & \longrightarrow & M' & \longrightarrow & I^0 & \longrightarrow & I^1 & \longrightarrow & I^2 & \longrightarrow & \cdots \end{array}$$

Suppose that the top row is exact, and that each I^n ($n \geqq 0$) is injective. Let $\varphi : M \to M'$ be a given homomorphism. Then there exists a morphism f of complexes such that $f_{-1} = \varphi$; and any two such are homotopic.

Proof. By definition of an injective, the homomorphism $M \to I^0$ via M' extends to a homomorphism

$$f_0 : E^0 \to I^0,$$

which makes the first square commute:

$$\begin{array}{ccc} M & \longrightarrow & E_0 \\ \Big\downarrow{\scriptstyle \varphi} & & \Big\downarrow{\scriptstyle f_0} \\ M' & \longrightarrow & I^0 \end{array}$$

Next we must construct f_1. We write the second square in the form

$$\begin{array}{ccccc} 0 & \longrightarrow & E^0/M & \longrightarrow & E^1 \\ & & \big\downarrow{\scriptstyle f_0} & & \\ & & I^0 & \longrightarrow & I^1 \end{array}$$

with the exact top row as shown. Again because I^1 is injective, we can apply the same argument and find f_1 to make the second square commute. And so on, thus constructing the morphism of complexes f.

Suppose f, g are two such morphisms. We define $h_0 : E^0 \to M'$ to be 0. Then the condition for a homotopy is satisfied in the first instance, when

$$f_{-1} = g_{-1} = \varphi.$$

Next let $d^{-1} : M \to E^0$ be the embedding of M in E^0. Since I^0 is injective, we can extend

$$d^0 : E^0/\mathrm{Im}\, d^{-1} \to E_1$$

to a homomorphism $h_1 : E^1 \to I^0$. Then the homotopy condition is verified for $f_0 - g_0$. Since $h_0 = 0$ we actually have in this case

$$f_0 - g_0 = h_1 d^0,$$

but this simplification is misleading for the inductive step which follows. We assume constructed the map h_{n+1}, and we wish to show the existence of h_{n+2} satisfying

$$f_{n+1} - g_{n+1} = d'^n h_{n+1} + h_{n+2} d^{n+1}.$$

Since $\mathrm{Im}\, d^n = \mathrm{Ker}\, d^{n+1}$, we have a monomorphism $E^{n+1}/\mathrm{Im}\, d^n \to E^{n+2}$. By the definition of an injective object, which in this case is I^{n+1}, it suffices to prove that

$$f_{n+1} - g_{n+1} - d'^n h_{n+1} \quad \text{vanishes on the image of } d^n,$$

and to use the exact diagram:

$$\begin{array}{ccccc} 0 & \longrightarrow & E^{n+1}/\mathrm{Im}\, d^n & \longrightarrow & E^{n+2} \\ & & \big\downarrow{\scriptstyle f_{n+1} - g_{n+1}} & & \\ & & I^{n+1} & & \end{array}$$

to get the existence of $h_{n+2} : E^{n+2} \to I^{n+1}$ extending $f_{n+1} - g_{n+1}$. But we have:

$$\begin{aligned} &(f_{n+1} - g_{n+1} - d'^n h_{n+1}) d^n \\ &\quad = (f_{n+1} - g_{n+1}) d^n - d'^n h_{n+1} d^n \end{aligned}$$

$= (f_{n+1} - g_{n+1})d^n - d'^n(f_n - g_n - d'^{(n-1)}h_n)$ by induction

$= (f_{n+1} - g_{n+1})d^n - d'^n(f_n - g_n)$ because $d'd' = 0$

$= 0$ because f, g are homomorphisms of complexes.

This concludes the proof of Lemma 5.2.

Remark. Dually, let $P_{M'} \to M' \to 0$ be a complex with P^i projective for $i \geqq 0$, and let $E_M \to M \to 0$ be a resolution. Let $\varphi: M' \to M$ be a homomorphism. Then φ extends to a homomorphism of complex $P \to E$. The proof is obtained by reversing arrows in Lemma 5.2. The books on homological algebra that I know of in fact carry out the projective case, and leave the injectivc case to the reader. However, one of my motivations is to do here what is needed, for instance in [Ha 77], Chapter III, on derived functors, as a preliminary to the cohomology of sheaves. For an example of projective resolutions using free modules, see Exercises 2–7, concerning the cohomology of groups.

§6. DERIVED FUNCTORS

We continue to work in an abelian category. A covariant additive functor

$$F: \mathfrak{A} \to \mathfrak{B}$$

is said to be **left exact** if it transforms an exact sequence

$$0 \to M' \to M \to M''$$

into an exact sequence $0 \to F(M') \to F(M) \to F(M'')$. We remind the reader that F is called **additive** if the map

$$\text{Hom}(A', A) \to \text{Hom}(FA', FA)$$

is additive.

We assume throughout that F is left exact unless otherwise specified, and additive. We continue to assume that our abelian category has enough injectives.

Given an object M, let

$$0 \to M \to I^0 \to I^1 \to I^2 \to$$

be an injective resolution, which we abbreviate by

$$0 \to M \to I_M,$$

where I_M is the complex $I^0 \to I^1 \to I^2 \to$. We let I be the complex

$$0 \to I^0 \to I^1 \to I^2 \to$$

We define the **right-derived functor** R^nF by

$$R^nF(M) = H^n(F(I)),$$

in other words, the n-th homology of the complex

$$0 \to F(I^0) \to F(I^1) \to F(I^2) \to$$

Directly from the definitions and the monomorphism $M \to I_0$, we see that there is an isomorphism

$$R^0F(M) = F(M).$$

This isomorphism seems at first to depend on the injective resolution, and so do the functors $R^nF(M)$ for other n. However, from Lemmas 5.1 and 5.2 we see that given two injective resolutions of M, there is a homomorphism between them, and that any two homomorphisms are homotopic. If we apply the functor F to these homomorphisms and to the homotopy, then we see that the homology of the complex $F(I)$ is in fact determined up to a unique isomorphism. One therefore omits the resolution from the notation and from the language.

Example 1. Let R be a ring and let $\mathfrak{A} = \text{Mod}(R)$ be the category of R-modules. Fix a module A. The functor $M \mapsto \text{Hom}(A, M)$ is left exact, i.e. given an exact sequence $0 \to M' \to M \to M''$, the sequence

$$0 \to \text{Hom}(A, M') \to \text{Hom}(A, M) \to \text{Hom}(A, M'')$$

is exact. Its right derived functors are denoted by $\text{Ext}^n(A, M)$ for M variable. Similarly, for a fixed module B, the functor $X \mapsto \text{Hom}(X, B)$ is right exact, and it gives rise to its **left derived functors**. For the explicit mirror image of the terminology, see the end of this section. In any case, we may consider A as variable. In §8 we shall go more deeply into this aspect of the formalism, by dealing with bifunctors. It will turn out that $\text{Ext}^n(A, B)$ has a dual interpretation as a left derived functor of the first variable and right derived functor of the second variable. See Corollary 8.5.

In the exercises, you will prove that $\text{Ext}^1(A, M)$ is in bijection with isomorphism classes of extensions, of M by A, that is, isomorphism classes of exact sequences

$$0 \to A \to E \to M \to 0.$$

The name Ext comes from this interpretation in dimension 1.

For the computation of Ext^i in certain important cases, see Chapter XXI, Theorems 4.6 and 4.11, which serve as examples for the general theory.

Example 2. Let R be commutative. The functor $M \mapsto A \otimes M$ is right exact, in other words, the sequence

$$A \otimes M' \to A \otimes M \to A \otimes M'' \to 0$$

is exact. Its left derived functors are denoted by $\text{Tor}_n(A, M)$ for M variable.

Example 3. Let G be a group and let $R = \mathbf{Z}[G]$ be the group ring. Let $\mathfrak{A}$ be the category of G-modules, i.e. $\mathfrak{A} = \text{Mod}(R)$, also denoted by Mod($G$). For a G-module A, let A^G be the submodule (abelian group) consisting of those elements v such that $xv = v$ for all $x \in G$. Then $A \mapsto A^G$ is a left exact functor from Mod(R) into the category of abelian groups. Its left derived functors give rise to the cohomology of groups. Some results from this special cohomology will be carried out in the exercises, as further examples of the general theory.

Example 4. Let X be a topological space (we assume the reader knows what this is). By a **sheaf** $\mathfrak{F}$ of abelian groups on X, we mean the data:

(a) For every open set U of X there is given an abelian group $\mathfrak{F}(U)$.

(b) For every inclusion $V \subset U$ of open sets there is given a homomorphism

$$\text{res}_V^U : \mathfrak{F}(U) \to \mathfrak{F}(V),$$

called the **restriction** from U to V, subject to the following conditions:

SH 1. $\mathfrak{F}(\text{empty set}) = 0$.

SH 2. res_U^U is the identity $\mathfrak{F}(U) \to \mathfrak{F}(U)$.

SH 3. If $W \subset V \subset U$ are open sets, then $\text{res}_W^V \circ \text{res}_V^U = \text{res}_W^U$.

SH 4. Let U be an open set and $\{V_i\}$ be an open covering of U. Let $s \in \mathfrak{F}(U)$. If the restriction of s to each V_i is 0, then $s = 0$.

SH 5. Let U be an open set and let $\{V_i\}$ be an open covering of U. Suppose given $s_i \in \mathfrak{F}(V_i)$ for each i, such that given i, j the restrictions of s_i and s_j to $V_i \cap V_j$ are equal. Then there exists a unique $s \in \mathfrak{F}(U)$ whose restriction to V_i is s_i for all i.

Elements of $\mathfrak{F}(U)$ are called **sections** of $\mathfrak{F}$ over U. Elements of $\mathfrak{F}(X)$ are called **global sections**. Just as for abelian groups, it is possible to define the notion of homomorphisms of sheaves, kernels, cokernels, and exact sequences. The association $\mathfrak{F} \mapsto \mathfrak{F}(X)$ (global sections functor) is a functor from the category of sheaves of abelian groups to abelian groups, and this functor is left exact. Its right derived functors are the basis of cohomology theory in topology and algebraic geometry (among other fields of mathematics). The reader will find a self-contained brief definition of the basic properties in [Ha 77], Chapter II, §1, as well as a proof that these form an abelian category. For a more extensive treatment I recommend Gunning's [Gu 91], mentioned in the introduction to Part IV, notably Volume III, dealing with the cohomology of sheaves.

We now return to the general theory of derived functors. The general theory tells us that these derived functors do not depend on the resolution by projectives or injectives according to the variance. As we shall also see in §8, one can even use other special types of objects such as acyclic or exact (to be defined), which gives even more flexibility in the ways one has to compute homology. Through certain explicit resolutions, we obtain means of computing the derived functors

explicitly. For example, in Exercise 16, you will see that the cohomology of finite cyclic groups can be computed immediately by exhibiting a specific free resolution of **Z** adapted to such groups. Chapter XXI will contain several other examples which show how to construct explicit finite free resolutions, which allow the determination of derived functors in various contexts.

The next theorem summarizes the basic properties of derived functors.

Theorem 6.1. *Let $\mathfrak{A}$ be an abelian category with enough injectives, and let $F : \mathfrak{A} \to \mathfrak{B}$ be a covariant additive left exact functor to another abelian category $\mathfrak{B}$. Then:*

(i) *For each $n \geqq 0$, R^nF as defined above is an additive functor from $\mathfrak{A}$ to $\mathfrak{B}$. Furthermore, it is independent, up to a unique isomorphism of functors, of the choices of resolutions made.*

(ii) *There is a natural isomorphism $F \approx R^0F$.*

(iii) *For each short exact sequence*

$$0 \to M' \to M \to M'' \to 0$$

and for each $n \geqq 0$ there is a natural homomorphism

$$\delta^n : R^nF(M'') \to R^{n+1}F(M)$$

such that we obtain a long exact sequence:

$$\to R^nF(M') \to R^nF(M) \to R^nF(M'') \xrightarrow{\delta^n} R^{n+1}F(M') \to .$$

(iv) *Given a morphism of short exact sequences*

$$\begin{array}{ccccccccc} 0 & \longrightarrow & M' & \longrightarrow & M & \longrightarrow & M'' & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & N' & \longrightarrow & N & \longrightarrow & N'' & \longrightarrow & 0 \end{array}$$

the δ's give a commutative diagram:

$$\begin{array}{ccc} R^nF(M'') & \xrightarrow{\delta^n} & R^{n+1}F(M') \\ \downarrow & & \downarrow \\ R^nF(N'') & \xrightarrow[\delta^n]{} & R^{n+1}F(N') \end{array}$$

(v) *For each injective object I of A and for each $n > 0$ we have $R^nF(I) = 0$.*

Properties (i), (ii), (iii), and (iv) essentially say that R^nF is a delta-functor in a sense which will be expanded in the next section. The last property (v) will be discussed after we deal with the delta-functor part of the theorem.

We now describe how to construct the δ-homomorphisms. Given a short exact sequence, we can find an injective resolution of M', M, M'' separately, but they don't necessarily fit in an exact sequence of complexes. So we must achieve this to apply the considerations of §1. Consider the diagram:

$$\begin{array}{ccccccccc} & & 0 & & 0 & & 0 & & \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \to & M' & \to & M & \to & M'' & \to & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \to & I'^0 & \to & X & \to & I''^0 & \to & 0. \end{array}$$

We give monomorphisms $M' \to I'^0$ and $M'' \to I''^0$ into injectives, and we want to find X injective with a monomorphism $M \to X$ such that the diagram is exact. We take X to be the direct sum

$$X = I'^0 \oplus I''^0.$$

Since I'^0 is injective, the monomorphism $M' \to I'^0$ can be extended to a homomorphism $M \to I'^0$. We take the homomorphism of M into $I'^0 \oplus I''^0$ which comes from this extension on the first factor I'^0, and is the composite map

$$M \to M'' \to I''^0$$

on the second factor. Then $M \to X$ is a monomorphism. Furthermore $I'^0 \to X$ is the monomorphism on the first factor, and $X \to I''^0$ is the projection on the second factor. So we have constructed the diagram we wanted, giving the beginning of the compatible resolutions.

Now we take the quotient homomorphism, defining the third row, to get an exact diagram:

$$\begin{array}{ccccccccc} & & 0 & & 0 & & 0 & & \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \to & M' & \to & M & \to & M'' & \to & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \to & I'^0 & \to & I^0 & \to & I''^0 & \to & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \to & N' & \to & N & \to & N'' & \to & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ & & 0 & & 0 & & 0 & & \end{array}$$

where we let $I^0 = X$, and N', N, N'' are the cokernels of the vertical maps by definition. The exactness of the N-sequence is left as an exercise to the reader. We then repeat the construction with the N-sequence, and by induction construct injective resolutions

$$\begin{array}{ccccccccc} & & 0 & & 0 & & 0 & & \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & M' & \longrightarrow & M & \longrightarrow & M'' & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & I'_{M'} & \longrightarrow & I_M & \longrightarrow & I''_{M''} & \longrightarrow & 0 \end{array}$$

of the M-sequence such that the diagram of the resolutions is exact.

We now apply the functor F to this diagram. We obtain a short sequence of complexes:

$$0 \to F(I') \to F(I) \to F(I'') \to 0,$$

which is exact because $I = I' \oplus I''$ is a direct sum and F is left exact, so F commutes with direct sums. We are now in a position to apply the construction of §1 to get the coboundary operator in the homology sequence:

$$R^nF(M') \to R^nF(M) \to R^nF(M'') \xrightarrow{\delta^n} R^{n+1}F(M').$$

This is legitimate because the right derived functor is independent of the chosen resolutions.

So far, we have proved (i), (ii), and (iii). To prove (iv), that is the naturality of the delta homomorphisms, it is necessary to go through a three-dimensional commutative diagram. At this point, I feel it is best to leave this to the reader, since it is just more of the same routine.

Finally, the last property (v) is obvious, for if I is injective, then we can use the resolution

$$0 \to I \to I \to 0$$

to compute the derived functors, from which it is clear that $R^nF = 0$ for $n > 0$.

This concludes the proof of Theorem 6.1.

In applications, it is useful to determine the derived functors by means of other resolutions besides injective ones (which are useful for theoretical purposes, but not for computational ones). Let again F be a left exact additive functor. An object X is called **F-acyclic** if $R^nF(X) = 0$ for all $n > 0$.

Theorem 6.2. *Let*

$$0 \to M \to X^0 \to X^1 \to X^2 \to \cdots$$

be a resolution of M by F-acyclics. Let

$$0 \to M \to I^0 \to I^1 \to I^2 \to \cdots$$

be an injective resolution. Then there exists a morphism of complexes $X_M \to I_M$ extending the identity on M, and this morphism induces an isomorphism

$$H^nF(X) \approx H^nF(I) = R^nF(M) \qquad \text{for all } n \geqq 0.$$

Proof. The existence of the morphism of complexes extending the identity on M is merely Lemma 5.2. The usual proof of the theorem via spectral sequences can be formulated independently in the following manner, shown to me by David Benson. We need a lemma.

Lemma 6.3. *Let Y^i ($i \geqq 0$) be F-acyclic, and suppose the sequence*

$$0 \to Y^0 \to Y^1 \to Y^2 \to \cdots$$

is exact. Then

$$0 \to F(Y^0) \to F(Y^1) \to F(Y^2) \to \cdots$$

is exact.

Proof. Since F is left exact, we have an exact sequence

$$0 \to F(Y^0) \to F(Y^1) \to F(Y^2).$$

We want to show exactness at the next joint. We draw the cokernels:

$$\begin{array}{ccccccccc}
0 & \longrightarrow & Y^0 & \longrightarrow & Y^1 & \longrightarrow & Y^2 & \longrightarrow & Y^3 \\
 & & & & & \searrow & \nearrow & \searrow & \nearrow \\
 & & & & & Z^1 & & Z^2 & \\
 & & & & \nearrow & \searrow & \nearrow & \searrow & \\
 & & & & 0 & & 0 & & 0
\end{array}$$

So $Z_1 = \operatorname{Coker}(Y^0 \to Y^1)$; $Z_2 = \operatorname{Coker}(Y^1 \to Y^2)$; etc. Applying F we have an exact sequence

$$0 \to F(Y^0) \to F(Y^1) \to F(Z^1) \to R^1F(Y^0) = 0.$$

So $F(Z_1) = \mathrm{Coker}(F(Y^0) \to F(Y^1))$. We now consider the exact sequence

$$0 \to Z_1 \to Y_2 \to Y_3$$

giving the exact sequence

$$0 \to F(Z^1) \to F(Y^2) \to F(Y^3)$$

by the left-exactness of F, and proving what we wanted. But we can now continue by induction because Z_1 is F-acyclic, by the exact sequence

$$0 \to R^nF(Y^1) \to R^nF(Z^1) \to R^{n+1}F(Y^0) = 0.$$

This concludes the proof of Lemma 6.3.

We return to the proof of Theorem 6.2. The injective resolution

$$0 \to M \to I_M$$

can be chosen such that the homomorphisms $X_n \to I_n$ are monomorphisms for $n \geqq 0$, because the derived functor is independent of the choice of injective resolution. Thus we may assume without loss of generality that we have an exact diagram:

$$\begin{array}{ccccccccccc}
 & & & & 0 & & 0 & & 0 & & \\
 & & & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & M & \longrightarrow & X^0 & \longrightarrow & X^1 & \longrightarrow & X^2 & \longrightarrow & \\
 & & {\scriptstyle \mathrm{id}}\downarrow & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & M & \longrightarrow & I^0 & \longrightarrow & I^1 & \longrightarrow & I^2 & \longrightarrow & \\
 & & & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & \longrightarrow & Y^0 & \longrightarrow & Y^1 & \longrightarrow & Y^2 & \longrightarrow & \\
 & & & & \downarrow & & \downarrow & & \downarrow & & \\
 & & & & 0 & & 0 & & 0 & &
\end{array}$$

defining Y^n as the appropriate cokernel of the vertical map.

Since X^n and I^n are acyclic, so is Y^n from the exact sequence

$$R^kF(I^n) \to R^kF(Y^n) \to R^{k+1}F(X^n).$$

Applying F we obtain a short exact sequence of complexes

$$0 \to F(X) \to F(I) \to F(Y) \to 0.$$

whence the corresponding homology sequence

$$H^{n-1}F(Y) \to H^nF(X) \to H^nF(I) \to H^nF(Y).$$

Both extremes are 0 by Lemma 6.3, so we get an isomorphism in the middle, which by definition is the isomorphism

$$H^nF(X) \approx R^nF(M),$$

thus proving the theorem.

Left derived functors

We conclude this section by a summary of the properties of left derived functors.

We consider complexes going the other way,

$$\to X_n \to \cdots \to X_2 \to X_1 \to X_0 \to M \to 0$$

which we abbreviate by

$$X_M \to M \to 0.$$

We call such a complex a **resolution** of M if the sequence is exact. We call it a **projective resolution** if X_n is projective for all $n \geqq 0$.

Given projective resolutions X_M, $Y_{M'}$ and a homomorphism

$$\varphi : M \to M'$$

there always exists a homomorphism $X_M \to Y_{M'}$ extending φ, and any two such are homotopic.

In fact, one need only assume that X_M is a projective resolution, and that $Y_{M'}$ is a resolution, not necessarily projective, for the proof to go through.

Let T be a covariant additive functor. Fix a projective resolution of an object M,

$$P_M \to M \to 0.$$

We define the **left derived functor** L_nT by

$$L_nT(M) = H_n(T(P)),$$

where $T(P)$ is the complex

$$\to T(P_n) \to \cdots \to T(P_2) \to T(P_1) \to T(P_0) \to 0.$$

The existence of homotopies shows that $L_nT(M)$ is uniquely determined up to a unique isomorphism if one changes the projective resolution.

We define T to be **right exact** if an exact sequence

$$M' \to M \to M'' \to 0$$

yields an exact sequence

$$T(M') \to T(M) \to T(M'') \to 0.$$

If T is right exact, then we have immediately from the definitions

$$L_0 T(M) \approx M.$$

Theorems 6.1 and 6.2 then go over to this case with similar proofs. One has to replace "injectives" by "projectives" throughout, and in Theorem 6.1, the last condition states that for $n > 0$,

$$L_n T(P) = 0 \qquad \text{if } P \text{ is projective.}$$

Otherwise, it is just a question of reversing certain arrows in the proofs. For an example of such left derived functors, see Exercises 2–7 concerning the cohomology of groups.

§7. DELTA-FUNCTORS

In this section, we axiomatize the properties stated in Theorem 6.1 following Grothendieck.

Let $\mathfrak{A}$, $\mathfrak{B}$ be abelian categories. A (covariant) **δ-functor** from $\mathfrak{A}$ to $\mathfrak{B}$ is a family of additive functors $F = \{F_n\}_{n \geqq 0}$, and to each short exact sequence

$$0 \to M' \to M \to M'' \to 0$$

an associated family of morphisms

$$\delta^n \colon F^n(M'') \to F^{n+1}(M')$$

with $n \geqq 0$, satisfying the following conditions:

DEL 1. For each short exact sequence as above, there is a long exact sequence

$$0 \to F^0(M') \to F^0(M) \to F^0(M'') \to F^1(M') \to \cdots$$

$$\to F^n(M') \to F^n(M) \to F^n(M'') \to F^{n+1}(M') \to$$

DEL 2. For each morphism of one short exact sequence as above into another $0 \to N' \to N \to N'' \to 0$, the δ's give a commutative diagram:

$$\begin{array}{ccc} F^n(M'') & \xrightarrow{\delta} & F^{n+1}(M') \\ \downarrow & & \downarrow \\ F^n(N'') & \xrightarrow[\delta]{} & F^{n+1}(N'). \end{array}$$

Before going any further, it is useful to give another definition. Many proofs in homology theory are given by induction from one index to the next. It turns out that the only relevant data for going up by one index is given in two successive dimensions, and that the other indices are irrelevant. Therefore we generalize the notion of δ-functor as follows.

A **δ-functor defined in degrees 0, 1** is a pair of functors (F^0, F^1) and to each short exact sequence

$$0 \to A' \to A \to A'' \to 0$$

an associated morphism

$$\delta : F^0(A'') \to F^1(A'')$$

satisfying the two conditions as before, but putting $n = 0$, $n + 1 = 1$, and forgetting about all other integers n. We could also use any two consecutive positive integers to index the δ-functor, or any sequence of consecutive integers $\geqq 0$. In practice, only the case of all integers $\geqq 0$ occurs, but for proofs, it is useful to have the flexibility provided by using only two indices, say 0, 1.

The δ-functor F is said to be **universal**, if given any other δ-functor G of $\mathfrak{A}$ into $\mathfrak{B}$, and given any morphism of functors

$$f_0 : F^0 \to G^0,$$

there exists a unique sequence of morphisms

$$f_n : F^n \to G^n$$

for all $n \geqq 0$, which commute with the δ^n for each short exact sequence.

By the definition of universality, a δ-functor G such that $G^0 = F^0$ is uniquely determined up to a unique isomorphism of functors. We shall give a condition for a functor to be universal.

An additive functor F of $\mathfrak{A}$ into $\mathfrak{B}$ is called **erasable** if to each object A there exists a monomorphism $u : A \to M$ for some M such that $F(u) = 0$. In practice, it even happens that $F(M) = 0$, but we don't need it in the axiomatization.

Linguistic note. Grothendieck originally called the notion "effaceable" in French. The dictionary translation is "erasable," as I have used above. Apparently people who did not know French have used the French word in English, but there is no need for this, since the English word is equally meaningful and convenient.

We say the functor is erasable by **injectives** if in addition M can be taken to be injective.

Example. Of course, a right derived functor is erasable by injectives, and a left derived functor by projectives. However, there are many cases when one wants erasability by other types of objects. In Exercises 9 and 14, dealing with the cohomology of groups, you will see how one erases the cohomology functor with induced modules, or regular modules when G is finite. In the category of coherent sheaves in algebraic geometry, one erases the cohomology with locally free sheaves of finite rank.

Theorem 7.1. *Let $F = \{F^n\}$ be a covariant δ-functor from $\mathfrak{A}$ into $\mathfrak{B}$. If F^n is erasable for each $n > 0$, then F is universal.*

Proof. Given an object A, we erase it with a monomorphism u, and get a short exact sequence:

$$0 \to A \xrightarrow{\varphi} M \to X \to 0.$$

Let G be another δ-functor with given $f_0 \colon F^0 \to G^0$. We have an exact commutative diagram

$$\begin{array}{ccccccc} F^0(M) & \longrightarrow & F^0(X) & \xrightarrow{\delta'} & F^1(A) & \longrightarrow & 0 \\ \downarrow{\scriptstyle f_0} & & \downarrow{\scriptstyle f_0} & & \vdots{\scriptstyle f_1?} & & \\ G^0(M) & \longrightarrow & G^0(X) & \xrightarrow[\delta_G]{} & G^1(A) & & \end{array}$$

We get the 0 on the top right because of the erasability assumption that

$$F^1(\varphi) = 0.$$

We want to construct

$$f_1(A) \colon F^1(A) \to G^1(A)$$

which makes the diagram commutative, is functorial in A, and also commutes with the δ. Commutativity in the left square shows that Ker δ_F is contained in the kernel of $\delta_G \circ f_0$. Hence there exists a unique homomorphism

$$f_1(A) \colon F^1(A) \to G^1(A)$$

which makes the right square commutative. We are going to show that $f_1(A)$ satisfies the desired conditions. The rest of the proof then proceeds by induction following the same pattern.

We first prove the functoriality in A.

Let $u : A \to B$ be a morphism. We form the push-out P in the diagram

$$\begin{array}{ccc} A & \xrightarrow{\varphi} & M \\ \downarrow{\scriptstyle u} & & \downarrow \\ B & \longrightarrow & P \end{array}$$

Since φ is a monomorphism, it follows that $B \to P$ is a monomorphism also. Then we let $P \to N$ be a monomorphism which erases F_1. This yields a commutative diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & A & \longrightarrow & M & \longrightarrow & X & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle u} & & \downarrow{\scriptstyle v} & & \downarrow{\scriptstyle w} & & \\ 0 & \longrightarrow & B & \longrightarrow & N & \longrightarrow & Y & \longrightarrow & 0 \end{array}$$

where $B \to N$ is the composite $B \to P \to N$, and Y is defined to be the cokernel of $B \to N$.

Functoriality in A means that the following diagram is commutative.

$$\begin{array}{ccc} F^1(A) & \xrightarrow{F^1(u)} & F^1(B) \\ \downarrow{\scriptstyle f_1(A)} & & \downarrow{\scriptstyle f_1(B)} \\ G^1(A) & \xrightarrow[F^1(u)]{} & G^1(B) \end{array}$$

This square is the right-hand side of the following cube:

$$\begin{array}{l} F^0(X) \xrightarrow{\delta_F} F^1(A), \quad F^0(X) \xrightarrow{F^0(w)} F^0(Y), \quad F^1(A) \xrightarrow{F^1(u)} F^1(B), \quad F^0(Y) \xrightarrow{\delta_F} F^1(B) \\ F^0(X) \xrightarrow{f_0(X)} G^0(X), \quad F^0(Y) \to G^0(Y), \quad F^1(A) \to G^1(A), \quad F^1(B) \xrightarrow{f_1(B)} G^1(B) \\ G^0(X) \xrightarrow{\delta_G} G^1(A), \quad G^0(X) \xrightarrow{G^0(w)} G^0(Y), \quad G^1(A) \xrightarrow{G^1(u)} G^1(B), \quad G^0(Y) \xrightarrow{\delta_G} G^1(B) \end{array}$$

All the faces of the cube are commutative except possibly the right-hand face. It is then a general fact that if the top maps here denoted by δ_F are epimorphisms,

then the right-hand side is commutative also. This can be seen as follows. We start with $f_1(B)F^1(u)\delta_F$. We then use commutativity on the top of the cube, then the front face, then the left face, then the bottom, and finally the back face. This yields

$$f_1(B)F^1(u)\delta_F = G^1(u)f_1(A)\delta_F.$$

Since δ_F is an epimorphism, we can cancel δ_F to get what we want.

Second, we have to show that f_1 commutes with δ. Let

$$0 \to A' \to A \to A'' \to 0$$

be a short exact sequence. The same push-out argument as before shows that there exists an erasing monomorphism $0 \to A' \to M$ and morphisms v, w making the following diagram commutative:

$$\begin{array}{ccccccccc}
0 & \longrightarrow & A' & \longrightarrow & A & \longrightarrow & A'' & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle \mathrm{id}} & & \downarrow{\scriptstyle v} & & \downarrow{\scriptstyle w} & & \\
0 & \longrightarrow & A' & \longrightarrow & M & \longrightarrow & X & \longrightarrow & 0
\end{array}$$

Here X is defined as the appropriate cokernel of the bottom row. We now consider the following diagram:

$$\begin{array}{ccccc}
 & & F^0(A'') & & \\
 & \swarrow{\scriptstyle F^0(w)} & \downarrow{\scriptstyle f_0} & \searrow{\scriptstyle \delta_F} & \\
 & & G^0(A'') & & \\
F^0(X) & & \xrightarrow{\quad \delta_F \quad} & & F^1(A') \\
\downarrow{\scriptstyle f_0} & \swarrow{\scriptstyle G^0(w)} & & \searrow{\scriptstyle \delta_G} & \downarrow{\scriptstyle f_1(A')} \\
G^0(X) & & \xrightarrow{\quad \delta_G \quad} & & G^1(A')
\end{array}$$

Our purpose is to prove that the right-hand face is commutative. The triangles on top and bottom are commutative by the definition of a δ-functor. The

left-hand square is commutative by the hypothesis that f_0 is a morphism of functors. The front square is commutative by the definition of $f_1(A')$. Therefore we find:

$$\begin{aligned} f_1(A')\delta_F &= f_1(A')\delta_F F^0(w) && \text{(top triangle)} \\ &= \delta_F f_0 F^0(w) && \text{(front square)} \\ &= \delta_F G^0(w) f_0 && \text{(left square)} \\ &= \delta_F f_0 && \text{(bottom triangle).} \end{aligned}$$

This concludes the proof of Theorem 7.1, since instead of the pair of indices (0, 1) we could have used $(n, n + 1)$.

Remark. The morphism f_1 constructed in Theorem 7.1 depends functorially on f_0 in the following sense. Suppose we have three delta functors F, G, H defined in degrees 0, 1. Suppose given morphisms

$$f_0 : F^0 \to G^0 \quad \text{and} \quad g_0 : G^0 \to H^0.$$

Suppose that the erasing monomorphisms erase both F and G. Then we can construct f_1 and g_1 by applying the theorem. On the other hand, the composite

$$g_0 f_0 = h_0 : F^0 \to H^0$$

is also a morphism of functors, and the theorem yields the existence of a morphism

$$h_1 : F^1 \to H^1$$

such that (h_0, h_1) is a δ-morphism. By uniqueness, we therefore have

$$h_1 = g_1 f_1.$$

This is what we mean by the functorial dependence as mentioned above.

Corollary 7.2. *Assume that $\mathfrak{A}$ has enough injectives. Then for any left exact functor $F : \mathfrak{A} \to \mathfrak{B}$, the derived functors R^nF with $n \geqq 0$ form a universal δ-functor with $F \approx R^0F$, which is erasable by injectives. Conversely, if $G = \{G^n\}_{n \geqq 0}$ is a universal δ-functor, then G^0 is left exact, and the G^n are isomorphic to R^nG^0 for each $n \geqq 0$.*

Proof. If F is a left exact functor, then the $\{R^nF\}_{n \geqq 0}$ form a δ-functor by Theorem 6.1. Furthermore, for any object A, let $u : A \to I$ be a monomorphism of A into an injective. Then $R^nF(I) = 0$ for $n > 0$ by Theorem 6.1(iv), so $R^nF(u) = 0$. Hence R^nF is erasable for all $n > 0$, and we can apply Theorem 7.1.

Remark. As usual, Theorem 7.1 applies to functors with different variance. Suppose $\{F^n\}$ is a family of contravariant additive functors, with n ranging over

a sequence of consecutive integers, say for simplicity $n \geqq 0$. We say that F is a **contravariant δ-functor** if given an exact sequence

$$0 \to M' \to M \to M'' \to 0$$

then there is an associated family of morphisms

$$\delta^n : F^n(M') \to F^{n+1}(M')$$

satisfying **DEL 1** and **DEL 2** with M' interchanged with M'' and N' interchanged with N''. We say that F is **coerasable** if to each object A there exists an epimorphism $u : M \to A$ such that $F(u) = 0$. We say that F is **universal** if given any other δ-functor G of $\mathfrak{A}$ into $\mathfrak{B}$ and given a morphism of functors

$$f_0 : F^0 \to G^0$$

there exists a unique sequence of morphisms

$$f_n : F^n \to G^n$$

for all $n \geqq 0$ which commute with δ for each short exact sequence.

Theorem 7.1′. *Let $F = \{F^n\}$ (n ranging over a consecutive sequence of integers $\geqq 0$) be a contravariant δ-functor from $\mathfrak{A}$ into $\mathfrak{B}$, and assume that F^n is coerasable for $n \geqq 1$. Then F is universal.*

Examples of δ-functors with the variances as in Theorems 7.1 *and* 7.1′ will be given in the next section in connection with bifunctors.

Dimension shifting

Let $F = \{F^n\}$ be a contravariant delta functor with $n \geqq 0$. Let $\mathcal{E}$ be a family of objects which erases F^n for all $n \geqq 1$, that is $F^n(E) = 0$ for $n \geqq 1$ and $E \in \mathcal{E}$. Then such a family allows us to do what is called **dimension shifting** as follows. Given an exact sequence

$$0 \to Q \to E \to M \to 0$$

with $E \in \mathcal{E}$, we get for $n \geqq 1$ an exact sequence

$$0 = F^n(E) \to F^n(Q) \to F^{n+1}(M) \to F^{n+1}(E) = 0,$$

and therefore an isomorphism

$$F^n(Q) \xrightarrow{\approx} F^{n+1}(M),$$

which exhibits a shift of dimensions by one. More generally:

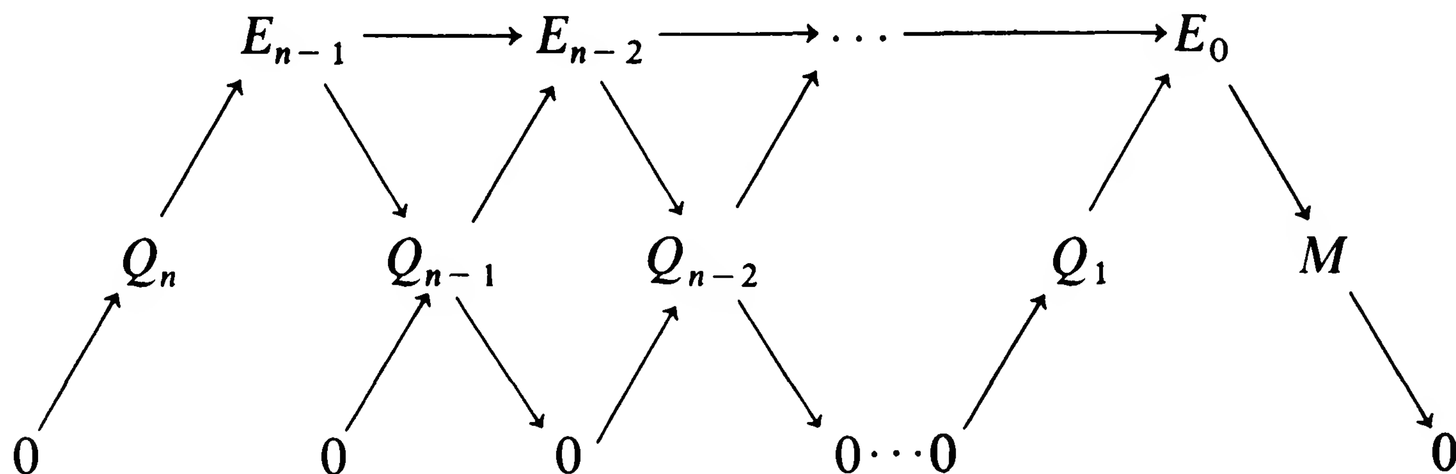
Proposition 7.3. *Let*

$$0 \to Q \to E_{n-1} \to \cdots \to E_0 \to M \to 0$$

be an exact sequence, such that $E_i \in \mathcal{E}$. Then we have an isomorphism

$$F^p(Q) \approx F^{p+n}(M) \qquad \textit{for } p \geqq 1.$$

Proof. Let $Q = Q_n$. Also without loss of generality, take $p = 1$. We may insert kernels and cokernels at each step as follows:



Then shifting dimension with respect to each short exact sequence, we find isomorphisms

$$F^1(Q_n) \approx F^2(Q_{n-1}) \approx \cdots \approx F^{n+1}(M).$$

This concludes the proof.

One says that M has F-**dimension** $\leqq d$ if $F^n(M) = 0$ for $n \geqq d + 1$. By dimension shifting, we see that if M has F-dimension $\leqq d$, then Q has F-dimension $\leqq d - n$ in Proposition 7.3. In particular, if M has F-dimension n, then Q has F-dimension 0.

The reader should rewrite all this formalism by changing notation, using for F the standard functors arising from Hom in the first variable, on the category of modules over a ring, which has enough projectives to erase the left derived functors of

$$A \mapsto \operatorname{Hom}(A, B),$$

for B fixed. We shall study this situation, suitably axiomatized, in the next section.

§8. BIFUNCTORS

In an abelian category one often deals with Hom, which can be viewed as a functor in two variables; and also the tensor product, which is a functor in two variables, but their variance is different. In any case, these examples lead to the notion of **bifunctor**. This is an association

$$(A, B) \mapsto T(A, B)$$

where A, B are objects of abelian categories $\mathfrak{A}$ and $\mathfrak{B}$ respectively, with values in some abelian category. This means that T is functorial in each variable, with the appropriate variance (there are four possibilities, with covariance and contravariance in all possible combinations); and if, say, T is covariant in all variables, we also require that for homomorphisms $A' \to A$ and $B' \to B$ there is a commutative diagram

$$\begin{array}{ccc} T(A', B') & \longrightarrow & T(A', B) \\ \downarrow & & \downarrow \\ T(A, B') & \longrightarrow & T(A, B). \end{array}$$

If the variances are shuffled, then the arrows in the diagram are to be reversed in the appropriate manner. Finally, we require that as a functor in each variable, T is additive.

Note that Hom is a bifunctor, contravariant in the first variable and covariant in the second. The tensor product is covariant in each variable.

The Hom functor is a bifunctor T satisfying the following properties:

HOM 1. *T is contravariant and left exact in the first variable.*

HOM 2. *T is covariant and left exact in the second variable.*

HOM 3. *For any injective object J the functor*

$$A \mapsto T(A, J)$$

is exact.

They are the only properties which will enter into consideration in this section. There is a possible fourth one which might come in other times:

HOM 4. *For any projective object Q the functor*

$$B \mapsto T(Q, B)$$

is exact.

But we shall deal *non-symmetrically, and view T as a functor of the second variable, keeping the first one fixed, in order to get derived functors of the second variable.* On the other hand, we shall also obtain a δ-functor of the first variable by using the bifunctor, even though this δ-functor is not a derived functor.

If $\mathfrak{B}$ has enough injectives, then we may form the right derived functors with respect to the second variable

$$B \mapsto R^n T(A, B), \qquad \text{also denoted by } R^n T_A(B),$$

fixing A, and viewing B as variable. If $T = \text{Hom}$, then this right derived functor is called **Ext**, so we have by definition

$$\text{Ext}^n(A, X) = R^n \text{Hom}(A, X).$$

We shall now give a criterion to compute the right derived functors in terms of the other (first) variable. We say that an object A is ***T*-exact** if the functor $B \mapsto T(A, B)$ is exact. By a ***T*-exact resolution** of an object A, we mean a resolution

$$\to M_1 \to M_0 \to A \to 0$$

where M_n is T-exact for all $n \geqq 0$.

Examples. Let $\mathfrak{A}$ and $\mathfrak{B}$ be the categories of modules over a commutative ring. Let $T = \text{Hom}$. Then a T-exact object is by definition a projective module. Now let the **transpose** of T be given by

$$^tT(A, B) = T(B, A).$$

Then a ^tT-exact object is by definition an injective module.

If T is the tensor product, such that $T(A, B) = A \otimes B$, then a T-exact object is called **flat**.

Remark. In the category of modules over a ring, there are enough projectives and injectives. But there are other situations when this is not the case. Readers who want to see all this abstract nonsense in action may consult [GriH 78], [Ha 77], not to speak of [SGA 6] and Grothendieck's collected works. It may genuinely happen in practice that $\mathfrak{B}$ has enough injectives but $\mathfrak{A}$ does not have enough projectives, so the situation is not all symmetric. Thus the functor $A \mapsto R^nT(A, B)$ for fixed B is *not* a derived functor in the variable A. In the above references, we may take for $\mathfrak{A}$ the category of coherent sheaves on a variety, and for $\mathfrak{B}$ the category of all sheaves. We let $T = \text{Hom}$. The locally free sheaves of finite rank are T-exact, and there are enough of them in $\mathfrak{A}$. There are enough injectives in $\mathfrak{B}$. And so it goes. The balancing act between T-exacts on one side, and injectives on the other is inherent to the situation.

Lemma 8.1. *Let T be a bifunctor satisfying* **HOM 1, HOM 2**. *Let $A \in \mathfrak{A}$, and let $M_A \to A \to 0$, that is*

$$\to M_1 \to M_0 \to A \to 0$$

be a T-exact resolution of A. Let $F^n(B) = H^n(T(M, B))$ for $B \in \mathfrak{B}$. Then F is a δ-functor and $F^0(B) = T(A, B)$. If in addition T satisfies **HOM 3**, *then $F^n(J) = 0$ for J injective and $n \geqq 1$.*

Proof. Given an exact sequence

$$0 \to B' \to B \to B'' \to 0$$

we get an exact sequence of complexes

$$0 \to T(M, B') \to T(M, B) \to T(M, B'') \to 0,$$

whence a cohomology sequence which makes F into a δ-functor. For $n = 0$ we get $F^0(B) = T(A, B)$ because $X \mapsto T(X, B)$ is contravariant and left exact for $X \in \mathfrak{A}$. If B is injective, then $F^n(B) = 0$ for $n \geqq 1$ by **HOM 3**, because $X \mapsto T(X, B)$ is exact. This proves the lemma.

Proposition 8.2. *Let T be a bifunctor satisfying* **HOM 1, HOM 2, HOM 3.** *Assume that $\mathfrak{B}$ has enough injectives. Let $A \in \mathfrak{A}$. Let*

$$M_A \to A \to 0$$

be a T-exact resolution of A. Then the two δ-functors

$$B \mapsto R^nT(A, B) \quad \text{and} \quad B \mapsto H^n(T(M, B))$$

are isomorphic as universal δ-functors vanishing on injectives, for $n \geqq 1$, and such that

$$R^0T(A, B) = H^0(T(M), B) = T(A, B).$$

Proof. This comes merely from the universality of a δ-functor erasable by injectives.

We now look at the functoriality in A.

Lemma 8.3. *Let T satisfy* **HOM 1, HOM 2,** *and* **HOM 3.** *Assume that $\mathfrak{B}$ has enough injectives. Let*

$$0 \to A' \to A \to A'' \to 0$$

be a short exact sequence. Then for fixed B, we have a long exact sequence

$$0 \to T(A'', B) \to T(A, B) \to T(A', B) \to$$
$$\to R^1T(A'', B) \to R^1T(A, B) \to R^1T(A', B) \to$$

such that the association

$$A \mapsto R^nT(A, B)$$

is a δ-functor.

Proof. Let $0 \to B \to I_B$ be an injective resolution of B. From the exactness of the functor $A \mapsto T(A, J)$, for J injective we get a short exact sequence of complexes

$$0 \to T(A'', I_B) \to T(A, I_B) \to T(A', I_B) \to 0.$$

Taking the associated long exact sequence of homology groups of these complexes yields the sequence of the proposition. (The functorality is left to the readers.)

If $T = \text{Hom}$, then the exact sequence looks like

$$0 \to \text{Hom}(A'', B) \to \text{Hom}(A, B) \to \text{Hom}(A', B) \to$$

$$\to \text{Ext}^1(A'', B) \to \text{Ext}^1(A, B) \to \text{Ext}^1(A', B) \to$$

and so forth.

We shall say that $\mathfrak{A}$ has **enough** T-exacts if given an object A in $\mathfrak{A}$ there is a T-exact M and an epimorphism

$$M \to A \to 0.$$

Proposition 8.4. *Let T satisfy* **HOM 1, HOM 2, HOM 3**. *Assume that $\mathfrak{B}$ has enough injectives. Fix $B \in \mathfrak{B}$. Then the association*

$$A \mapsto R^n T(A, B)$$

is a contravariant δ-functor on $\mathfrak{A}$ which vanishes on T-exacts, for $n \geqq 1$. If $\mathfrak{A}$ has enough T-exacts, then this functor is universal, coerasable by T-exacts, with value

$$R^0 T(A, B) = T(A, B).$$

Proof. By Lemma 8.3 we know that the association is a δ-functor, and it vanishes on T-exacts by Lemma 8.1. The last statement is then merely an application of the universality of erasable δ-functors.

Corollary 8.5. *Let $\mathfrak{A} = \mathfrak{B}$ be the category of modules over a ring. For fixed B, let* $\text{ext}^n(A, B)$ *be the left derived functor of $A \mapsto \text{Hom}(A, B)$, obtained by means of projective resolutions of A. Then*

$$\text{ext}^n(A, B) = \text{Ext}^n(A, B).$$

Proof. Immediate from Proposition 8.4.

The following proposition characterizes T-exacts cohomologically.

Proposition 8.6. *Let T be a bifunctor satisfying* **HOM 1, HOM 2, HOM 3.** *Assume that $\mathfrak{B}$ has enough injectives. Then the following conditions are equivalent:*

TE 1. *A is T-exact.*

TE 2. *For every B and every integer $n \geqq 1$, we have* $R^nT(A, B) = 0$.

TE 3. *For every B we have* $R^1T(A, B) = 0$.

Proof. Let

$$0 \to B \to I^0 \to I^1 \to$$

be an injective resolution of B. By definition, $R^nT(A, B)$ is the n-th homology of the sequence

$$0 \to T(A, I^0) \to T(A, I^1) \to T(A, I^2) \to$$

If A is T-exact, then this sequence is exact for $n \geqq 1$, so the homology is 0 and **TE 1** implies **TE 2**. Trivially, **TE 2** implies **TE 3**. Finally assume **TE 3**. Given an exact sequence

$$0 \to B' \to B \to B'' \to 0,$$

we have the homology sequence

$$0 \to T(A, B') \to T(A, B) \to T(A, B'') \to R^1T(A, B') \to .$$

If $R^1T(A, B') = 0$, then by definition A is T-exact, thus proving the proposition.

We shall say that an object A has T-**dimension** $\leqq d$ if

$$R^nT(A, B) = 0 \qquad \text{for } n > d \text{ and all } B.$$

Then the proposition states in particular that *A is T-exact if and only if A has T-dimension* 0.

Proposition 8.7. *Let T satisfy* **HOM 1, HOM 2, HOM 3.** *Assume that $\mathfrak{B}$ has enough injectives. Suppose that an object A admits a resolution*

$$0 \to E_d \to E_{d-1} \to \cdots \to E_0 \to A \to 0$$

where $E_0, \ldots, E_d$ are T-exact. Then A has T-dimension $\leqq d$. Assume this is the case. Let

$$0 \to Q \to L_{d-1} \to \cdots \to L_0 \to A \to 0$$

be a resolution where $L_0, \ldots, L_{d-1}$ are T-exact. Then Q is T-exact also.

Proof. By dimension shifting we conclude that Q has T-dimension 0, whence Q is T-exact by Proposition 8.6.

Proposition 8.7, like others, is used in the context of modules over a ring. In that case, we can take $T = \text{Hom}$, and

$$R^n T(A, B) = \text{Ext}^n(A, B).$$

For A to have T-dimension $\leqq d$ means that

$$\text{Ext}^n(A, B) = 0 \qquad \text{for } n > d \text{ and all } B.$$

Instead of T-exact, one can then read projective in the proposition.

Let us formulate the analogous result for a bifunctor that will apply to the tensor product. Consider the following properties.

TEN 1. *T is covariant and right exact in the first variable.*

TEN 2. *T is covariant and right exact in the second variable.*

TEN 3. *For any projective object P the functor*

$$A \mapsto T(A, P)$$

is exact.

As for Hom, there is a possible fourth property which will play no role in this section:

TEN 4. *For any projective object Q the functor*

$$B \mapsto T(Q, B)$$

is exact.

Proposition 8.2′. *Let T be a bifunctor satisfying* **TEN 1, TEN 2, TEN 3.** *Assume that $\mathfrak{B}$ has enough projectives. Let $A \in \mathfrak{A}$. Let*

$$M_A \to A \to 0$$

be a T-exact resolution of A. Then the two δ-functors

$$B \mapsto L_n T(A, B) \quad \text{and} \quad B \mapsto H_n(T(M, B))$$

are isomorphic as universal δ-functors vanishing on projectives, and such that

$$L_0 T(A, B) = H_0(T(M), B) = T(A, B).$$

Lemma 8.3′. *Assume that T satisfies* **TEN 1, TEN 2, TEN 3.** *Assume that $\mathfrak{B}$ has enough projectives. Let*

$$0 \to A' \to A \to A'' \to 0$$

be a short exact sequence. Then for fixed B, we have a long exact sequence:

$$\to L_1T(A', B) \to L_1T(A, B) \to L_1T(A'', B) \to$$
$$\to T(A', B) \to T(A, B) \to T(A'', B) \to 0$$

which makes the association $A \mapsto L_nT(A, B)$ a δ-functor.

Proposition 8.4′. *Let T satisfy* **TEN 1, TEN 2, TEN 3**. *Assume that $\mathfrak{B}$ has enough projectives. Fix $B \in \mathfrak{B}$. Then the association*

$$A \mapsto L_nT(A, B)$$

is a contravariant δ-functor on $\mathfrak{A}$ which vanishes on T-exacts for $n \geqq 1$. If $\mathfrak{A}$ has enough T-exacts, then this functor is universal, coerasable by T-exacts, with the value

$$L_0T(A, B) = T(A, B).$$

Corollary 8.8. *If there is a bifunctorial isomorphism $T(A, B) \approx T(B, A)$, and if B is T-exact, then for all A, $L_nT(A, B) = 0$ for $n \geqq 1$. In short, T-exact implies acyclic.*

Proof. Let $M_A = P_A$ be a projective resolution in Proposition 8.2′. By hypotheses, $X \mapsto T(X, B)$ is exact so $H_n(T(P, B)) = 0$ for $n \geqq 1$; so the corollary is a consequence of the proposition.

The above corollary is formulated so as to apply to the tensor product.

Proposition 8.6′. *Let T be a bifunctor satisfying* **TEN 1, TEN 2, TEN 3**. *Assume that $\mathfrak{B}$ has enough projectives. Then the following conditions are equivalent:*

TE 1. *A is T-exact.*

TE 2. *For every B and every integer $n \geqq 1$ we have $L_nT(A, B) = 0$.*

TE 3. *For every B, we have $L_1T(A, B) = 0$.*

Proof. We repeat the proof of 8.6 so the reader can see the arrows pointing in different ways.

Let

$$\to Q_1 \to Q_0 \to B \to 0$$

be a projective resolution of B. By definition, $L_nT(A, B)$ is the n-th homology of the sequence

$$\to T(A, Q_1) \to T(A, Q_0) \to 0.$$

If A is T-exact, then this sequence is exact for $n \geqq 1$, so the homology is 0, and **TE 1** implies **TE 2**. Trivially, **TE 2** implies **TE 3**. Finally, assume **TE 3**. Given an exact sequence

$$0 \to B' \to B \to B'' \to 0$$

we have the homology sequence

$$\to L_1 T(A, B'') \to T(A, B') \to T(A, B) \to T(A, B'') \to 0.$$

If $L_1 T(A, B'')$ is 0, then by definition, A is T-exact, thus proving the proposition.

§9. SPECTRAL SEQUENCES

This section is included for convenience of reference, and has two purposes: first, to draw attention to an algebraic gadget which has wide applications in topology, differential geometry, and algebraic geometry, see Griffiths-Harris, [GrH 78]; second, to show that the basic description of this gadget in the context in which it occurs most frequently can be done in just a few pages.

In the applications mentioned above, one deals with a filtered complex (which we shall define later), and a complex may be viewed as a graded object, with a differential d of degree 1. To simplify the notation at first, we shall deal with filtered objects and omit the grading index from the notation. This index is irrelevant for the construction of the spectral sequence, for which we follow Godement.

So let F be an object with a differential (i.e. endomorphism) d such that $d^2 = 0$. We assume that F is **filtered**, that is that we have a sequence

$$F = F^0 \supset F^1 \supset F^2 \supset \cdots \supset F^n \supset F^{n+1} = \{0\},$$

and that $dF^p \subset F^p$. This data is called a **filtered differential object**. (We assume that the filtration ends with 0 after a finite number of steps for convenience.)

One defines the **associated graded object**

$$\operatorname{Gr} F = \bigoplus_{p \geqq 0} \operatorname{Gr}^p F \quad \text{where} \quad \operatorname{Gr}^p F = F^p/F^{p+1}.$$

In fact, $\operatorname{Gr} F$ is a complex, with a differential of degree 0 induced by d itself, and we have the homology $H(\operatorname{Gr}^p F)$.

The filtration $\{F^p\}$ also induces a filtration on the homology $H(F, d) = H(F)$; namely we let

$$H(F)^p = \text{image of } H(F^p) \text{ in } H(F).$$

Since d maps F^p into itself, $H(F^p)$ is the homology of F^p with respect to the restriction of d to F^p, and it has a natural image in $H(F)$ which yields this filtration. In particular, we then obtain a graded object associated with the filtered homology, namely

$$\operatorname{Gr} H(F) = \bigoplus \operatorname{Gr}^p H(F).$$

A **spectral sequence** is a sequence $\{E_r, d_r\}$ $(r \geqq 0)$ of graded objects

$$E_r = \bigoplus_{p \geqq 0} E_r^p$$

together with homomorphisms (also called **differentials**) of degree r,

$$d_r : E_r^p \to E_r^{p+r}$$

satisfying $d_r^2 = 0$, and such that the homology of E_r is E_{r+1}, that is

$$H(E_r) = E_{r+1}.$$

In practice, one usually has $E_r = E_{r+1} = \cdots$ for $r \geqq r_0$. This limit object is called E_∞, and one says that the spectral sequence **abuts** to E_∞. Actually, to be perfectly strict, instead of equalities one should really be given isomorphisms, but for simplicity, we use equalities.

Proposition 9.1. *Let F be a filtered differential object. Then there exists a spectral sequence $\{E_r\}$ with:*

$$E_0^p = F^p/F^{p+1}; \qquad E_1^p = H(\operatorname{Gr}^p F); \qquad E_\infty^p = \operatorname{Gr}^p H(F).$$

Proof. Define

$$\boxed{\begin{aligned} &Z_r^p = \{x \in F^p \text{ such that } dx \in F^{p+r}\} \\ &E_r^p = Z_r^p/[dZ_{r-1}^{p-(r-1)} + Z_{r-1}^{p+1}]. \end{aligned}}$$

The definition of E_r^p makes sense, since Z_r^p is immediately verified to contain $dZ_{r-1}^{p-(r-1)} + Z_{r-1}^{p+1}$. Furthermore, d maps Z_r^p into Z_r^{p+r}, and hence includes a homomorphism

$$d_r : E_r^p \to E_r^{p+r}.$$

We shall now compute the homology and show that it is what we want.

First, for the cycles: An element $x \in Z_r^p$ represents a cycle of degree p in E_r if and only if $dx \in dZ_{r-1}^{p+1} + Z_{r-1}^{p+r+1}$, in other words

$$dx = dy + z, \qquad \text{with } y \in Z_{r-1}^{p+1} \quad \text{and} \quad z \in Z_{r-1}^{p+r+1}.$$

Write $x = y + u$, so $du = z$. Then $u \in F^p$ and $du \in F^{p+r+1}$, that is $u \in Z_{r+1}^p$. It follows that

$$p\text{-cycles of } E_r = (Z_{r+1}^p + Z_{r-1}^{p+1})/(dZ_{r-1}^{p-r+1} + Z_{r-1}^{p+1}).$$

On the other hand, the p-boundaries in E_r are represented by elements of dZ_r^{p-r}, which contains dZ_{r-1}^{p-r+1}. Hence

$$p\text{-boundaries of } E_r = (dZ_r^{p-r} + Z_{r-1}^{p+1})/(dZ_{r-1}^{p-r+1} + Z_{r-1}^{p+1}).$$

Therefore

$$\begin{aligned} H^p(E_r) &= (Z_{r+1}^p + Z_{r-1}^{p+1})/(dZ_r^{p-r} + Z_{r-1}^{p+1}) \\ &= Z_{r+1}^p/(Z_{r+1}^p \cap (dZ_r^{p-r} + Z_{r-1}^{p+1})). \end{aligned}$$

Since

$$Z_{r+1}^p \supset dZ_r^{p-r} \quad \text{and} \quad Z_{r+1}^p \cap Z_{r-1}^{p+1} = Z_r^{p+1},$$

it follows that

$$H^p(E_r) = Z_{r+1}^p/(dZ_r^{p-r} + Z_r^{p+1}) = E_{r+1}^p,$$

thus proving the property of a spectral sequence.

Remarks. It is sometimes useful in applications to note the relation

$$dZ_{r-1}^{p-(r-1)} + Z_{r-1}^{p+1} = Z_r^p \cap (dF^{p-r+1} + F^{p+1}).$$

The verification is immediate, but Griffiths-Harris use the expression on the right in defining the spectral sequence, whereas Godement uses the expression on the left as we have done above. Thus the spectral sequence may also be defined by

$$\boxed{E_r^p = Z_r^p \mod(dF^{p-r+1} + F^{p+1}).}$$

This is to be interpreted in the sense that Z mod S means

$$(Z + S)/S \quad \text{or} \quad Z/(Z \cap S).$$

The term E_0^p is F^p/F^{p+1} immediately from the definitions, and by the general property already proved, we get $E_1^p = H(F^p/F^{p+1})$. As to E_∞^p, for r large we have $Z_r^p = Z^p =$ cycles in F^p, and

$$\boxed{E_\infty^p = Z^p/(Z^{p+1} + (dF^0 \cap F^p))}$$

which is independent of r, and is precisely $\mathrm{Gr}^p H(F)$, namely the p-graded component of $H(F)$, thus proving the theorem.

The differential d_1 can be specified as follows.

Proposition 9.2. *The homomorphism*

$$d_1 : E_1^p \to E_1^{p+1}$$

is the coboundary operator arising from the exact sequence

$$0 \to F^{p+1}/F^{p+2} \to F^p/F^{p+2} \to F^p/F^{p+1} \to 0$$

viewing each term as a complex with differential induced by d.

Proof. Indeed, the coboundary

$$\delta : E_1^p = H(F^p/F^{p+1}) \to H(F^{p+1}/F^{p+2}) = E_1^{p+1}$$

is defined on a representative cycle z by dz, which is the same way that we defined d_1.

In most applications, the filtered differential object is itself graded, because it arises from the following situation. Let K be a complex, $K = (K^p, d)$ with $p \geqq 0$ and d of degree 1. By a **filtration** FK, also called a **filtered complex**, we mean a decreasing sequence of subcomplexes

$$K = F^0K \supset F^1K \supset F^2K \supset \cdots \supset F^nK \supset F^{n+1}K = \{0\}.$$

Observe that a short exact sequence of complexes

$$0 \to K' \to K \to K'' \to 0$$

gives rise to a filtration $K \supset K' \supset \{0\}$, viewing K' as a subcomplex.

To each filtered complex FK we associated the complex

$$\mathrm{Gr}\, FK = \mathrm{Gr}\, K = \bigoplus_{p \geqq 0} \mathrm{Gr}^p\, K,$$

where

$$\mathrm{Gr}^p\, K = F^pK/F^{p+1}K,$$

and the differential is the obvious one. The filtration F^pK on K also induces a filtration $F^pH(K)$ on the cohomology, by

$$F^pH^q(K) = F^pZ^q/F^pB^q.$$

The associated graded homology is

$$\operatorname{Gr} H(K) = \bigoplus_{p,q} \operatorname{Gr}^p H^q(K),$$

where

$$\operatorname{Gr}^p H^q(K) = F^p H^q(K)/F^{p+1} H^q(K).$$

A **spectral sequence** is a sequence $\{E_r, d_r\}$ $(r \geqq 0)$ of bigraded objects

$$E_r = \bigoplus_{p,q \geqq 0} E_r^{p,q}$$

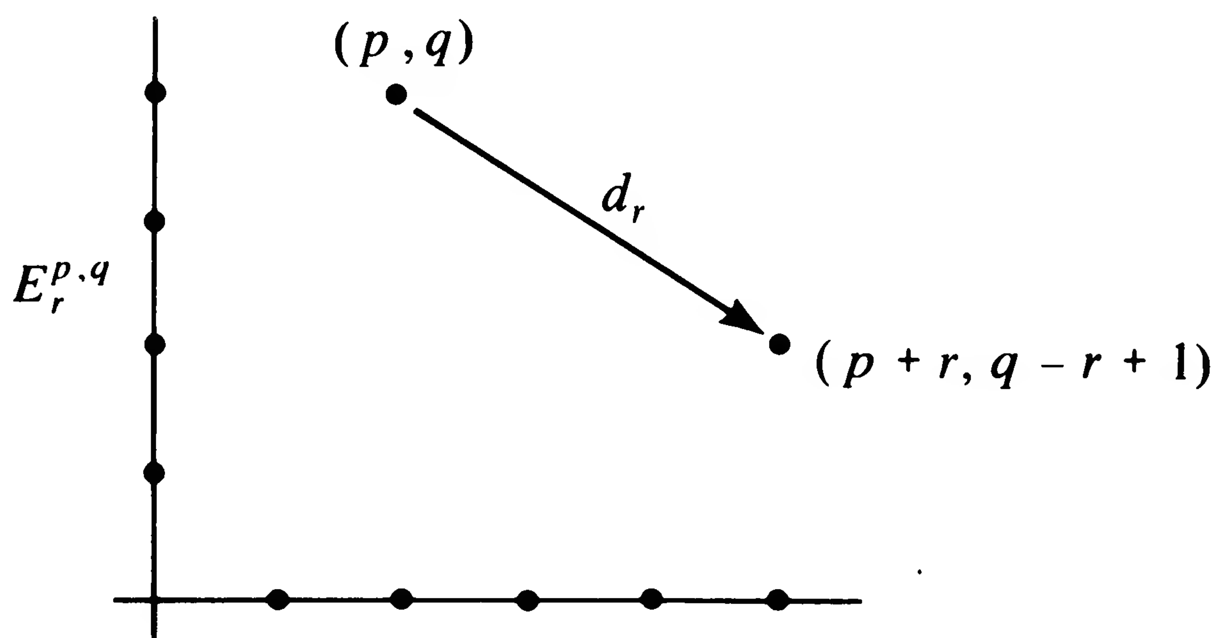
together with homomorphisms (called **differentials**)

$$d_r : E_r^{p,q} \to E_r^{p+r,q-r+1} \quad \text{satisfying} \quad d_r^2 = 0,$$

and such that the homology of E_r is E_{r+1}, that is

$$H(E_r) = E_{r+1}.$$

A spectral sequence is usually represented by the following picture:



In practice, one usually has $E_r = E_{r+1} = \cdots$ for $r \geqq r_0$. This limit object is called E_∞, and one says that the spectral sequence **abuts** to E_∞.

Proposition 9.3. *Let FK be a filtered complex. Then there exists a spectral sequence* $\{E_r\}$ *with:*

$$E_0^{p,q} = F^p K^{p+q}/F^{p+1} K^{p+q};$$

$$E_1^{p,q} = H^{p+q}(\operatorname{Gr}^p K);$$

$$E_\infty^{p,q} = \operatorname{Gr}^p (H^{p+q}(K)).$$

The last relation is usually written

$$E_r \Rightarrow H(K),$$

and we say that the spectral sequence **abuts** *to* $H(K)$.

The statement of Proposition 9.3 is merely a special case of Proposition 9.1, taking into account the extra graduation.

One of the main examples is the spectral sequence associated with a double complex

$$K = \bigoplus_{p,q \geqq 0} K^{p,q}$$

which is a bigraded object, together with differentials

$$d' : K^{p,q} \to K^{p+1,q} \quad \text{and} \quad d'' : K^{p,q} \to K^{p,q+1}$$

satisfying

$$d'^2 = d''^2 = 0 \quad \text{and} \quad d'd'' + d''d' = 0.$$

We denote the double complex by (K, d', d''). The associated single complex $(\mathrm{Tot}(K), D)$ (Tot for **total complex**), abbreviated K^*, is defined by

$$K^n = \bigoplus_{p+q=n} K^{p,q} \quad \text{and} \quad D = d' + d''.$$

There are two filtrations on (K^*, D) given by

$$'F^pK^n = \bigoplus_{\substack{p'+q=n \\ p' \geqq p}} K^{p',q}$$

$$''F^qK^n = \bigoplus_{\substack{p+q'=n \\ q' \geqq q}} K^{p,q'}.$$

There are two spectral sequences $\{'E_r\}$ and $\{''E_r\}$, both abutting to $H(\mathrm{Tot}(K))$. For applications, see [GrH 78], Chapter 3, §5; and also, for instance, [FuL 85], Chapter V. There are many situations when dealing with a double complex directly is a useful substitute for using spectral sequences, which are derived from double complexes anyhow.

We shall now derive the existence of a spectral sequence in one of the most important cases, the **Grothendieck spectral sequence** associated with the composite of two functors. *We assume that our abelian category has enough injectives.*

Let $C = \bigoplus C^p$ be a complex, and suppose $C^p = 0$ if $p < 0$ for simplicity. We define **injective resolution** of C to be a resolution

$$0 \to C \to I^0 \to I^1 \to I^2 \to \cdots$$

written briefly

$$0 \to C \to I_C$$

such that each I^j is a complex, $I^j = \bigoplus I^{j,p}$, with differentials

$$d^{j,p} : I^{j,p} \to I^{j,p+1}$$

and such that $I^{j,p}$ is an injective object. Then in particular, for each p we get an injective resolution of C^p, namely:

$$0 \to C^p \to I^{0,p} \to I^{1,p} \to \cdots$$

We let:

$$Z^{j,p} = \text{Ker } d^{j,p} = \text{cycles in degree } p$$
$$B^{j,p} = \text{Im } d^{j,p-1} = \text{boundaries in degree } p$$
$$H^{j,p} = Z^{j,p}/B^{j,p} = \text{homology in degree } p.$$

We then get complexes

$$0 \to Z^p(C) \to Z^{0,p} \to Z^{1,p} \to$$
$$0 \to B^p(C) \to B^{0,p} \to B^{1,p} \to$$
$$0 \to H^p(C) \to H^{0,p} \to H^{1,p} \to$$

We say that the resolution $0 \to C \to I_C$ is **fully injective** if these three complexes are injective resolutions of $Z^p(C)$, $B^p(C)$ and $H^p(C)$ respectively.

Lemma 9.4. *Let*

$$0 \to M' \to M \to M'' \to 0$$

be a short exact sequence. Let

$$0 \to M' \to I_{M'} \quad \textit{and} \quad 0 \to M'' \to I_{M''}$$

be injective resolutions of M' and M''. Then there exists an injective resolution

$$0 \to M \to I_M$$

of M and morphisms which make the following diagram exact and commutative:

$$\begin{array}{ccccccccc}
0 & \longrightarrow & I_{M'} & \longrightarrow & I_M & \longrightarrow & I_{M''} & \longrightarrow & 0 \\
 & & \uparrow & & \uparrow & & \uparrow & & \\
0 & \longrightarrow & M' & \longrightarrow & M & \longrightarrow & M'' & \longrightarrow & 0 \\
 & & \uparrow & & \uparrow & & \uparrow & & \\
 & & 0 & & 0 & & 0 & &
\end{array}$$

Proof. The proof is the same as at the beginning of the proof of Theorem 6.1.

Lemma 9.5. *Given a complex C there exists a fully injective resolution of C.*

Proof. We insert the kernels and cokernels in C, giving rise to the short exact sequences with boundaries B^p and cycles Z^p:

$$0 \to B^p \to Z^p \to H^p \to 0$$

$$0 \to Z^{p-1} \to C^{p-1} \to B^p \to 0.$$

We proceed inductively. We start with an injective resolution of

$$0 \to Z^{p-1} \to C^{p-1} \to B^p \to 0$$

using Lemma 9.4. Next let

$$0 \to H^p \to I_{H^p}$$

be an injective resolution of H^p. By Lemma 9.4 there exists an injective resolution

$$0 \to Z^p \to I_{Z^p}$$

which fits in the middle of the injective resolutions we already have for B^p and H^p. This establishes the inductive step, and concludes the proof.

Given a left exact functor G on an abelian category with enough injectives, we say that an object X is **G-acyclic** if $R^pG(X) = 0$ for $p \geqq 1$. Of course,

$$R^0G(X) = G(X).$$

Theorem 9.6. (Grothendieck spectral sequence). *Let*

$$T: \mathfrak{A} \to \mathfrak{B} \quad \textit{and} \quad G: \mathfrak{B} \to \mathfrak{C}$$

be covariant left exact functors such that if I is injective in $\mathfrak{A}$, then $T(I)$ is G-acyclic. Then for each A in $\mathfrak{A}$ there is a spectral sequence $\{E_r(A)\}$, such that

$$E_2^{p,q}(A) = R^pG(R^qT(A))$$

and $E_r^{p,q}$ abuts (with respect to p) to $R^{p+q}(GT)(A)$, where q is the grading index.

Proof. Let A be an object of $\mathfrak{A}$, and let $0 \to A \to C_A$ be an injective resolution. We apply T to get a complex

$$TC\colon \quad 0 \to TC^0 \to TC^1 \to TC^2 \to$$

By Lemma 9.5 there exists a fully injective resolution

$$0 \to TC \to I_{TC}$$

which has the 2-dimensional representation:

$$\begin{array}{ccccccccc}
 & & \uparrow & & \uparrow & & \uparrow & & \\
0 & \longrightarrow & I^{0,1} & \longrightarrow & I^{1,1} & \longrightarrow & I^{2,1} & \longrightarrow & \\
 & & \uparrow & & \uparrow & & \uparrow & & \\
0 & \longrightarrow & I^{0,0} & \longrightarrow & I^{1,0} & \longrightarrow & I^{2,0} & \longrightarrow & \\
 & & \uparrow & & \uparrow & & \uparrow & & \\
0 & \longrightarrow & TC^{0} & \longrightarrow & TC^{1} & \longrightarrow & TC^{2} & \longrightarrow & \\
 & & \uparrow & & \uparrow & & \uparrow & & \\
 & & 0 & & 0 & & 0 & &
\end{array}$$

Then GI is a double complex. Let $\mathrm{Tot}(GI)$ be the associated single complex. We now consider each of the two possible spectral sequences in succession, which we denote by ${}^1E_r^{p,q}$ and ${}^2E_r^{p,q}$.

The first one is the easiest. For fixed p, we have an injective resolution

$$0 \to TC^p \to I_{TC}^p$$

where we write I_{TC}^p instead of I_{TC^p}. This is the p-th column in the diagram. By definition of derived functors, GI^p is a complex whose homology is R^qG, in other words, taking homology with respect to d'' we have

$$''H^{p,q}(GI) = H^q(GI^p) = (R^qG)(TC^p).$$

By hypothesis, C^p injective implies that $(R^qG)(TC^p) = 0$ for $q > 0$. Since G is left exact, we have $R^0G(TC^p) = TC^p$. Hence we get

$$''H^{p,q}(GI) = \begin{cases} GT(C^p) & \text{if } q = 0 \\ 0 & \text{if } q > 0. \end{cases}$$

Hence the non-zero terms are on the p-axis, which looks like

$$0 \to GT(C^0) \to GT(C^1) \to GT(C^2) \to$$

Taking $'H^p$ we get

$${}^1E_2^{p,q}(A) = \begin{cases} R^p(GT)(A) & \text{if } q = 0 \\ 0 & \text{if } q > 0. \end{cases}$$

This yields

$$H^n(\mathrm{Tot}(GI)) \approx R^n(GT)(A).$$

The second one will use the full strength of Lemma 9.5, which had not been used in the first part of the proof, so it is now important that the resolution I_{TC} is fully injective. We therefore have injective resolutions

$$0 \to Z^p(TC) \to {}^1Z^{0,p} \to {}^1Z^{1,p} \to {}^1Z^{2,p} \to$$

$$0 \to B^p(TC) \to {}^1B^{0,p} \to {}^1B^{1,p} \to {}^1B^{2,p} \to$$

$$0 \to H^p(TC) \to {}^1H^{0,p} \to {}^1H^{1,p} \to {}^1H^{2,p} \to$$

and the exact sequences

$$0 \to {}^1Z^{q,p} \to I^{q,p} \to {}^1B^{q+1,p} \to 0$$

$$0 \to {}^1B^{q,p} \to {}^1Z^{q,p} \to {}^1H^{q,p} \to 0$$

split because of the injectivity of the terms. We denote by $I^{(p)}$ the p-th row of the double complex $I = \{I^{q,p}\}$. Then we find:

$$\begin{aligned} {}'H^{q,p}(GI) = H^q(GI^{(p)}) &= G^1Z^{q,p}/G^1B^{q,p} && \text{by the first split sequence} \\ &= G'H^{q,p}(I) && \text{by the second split sequence} \end{aligned}$$

because applying the functor G to a split exact sequence yields a split exact sequence.

Then

$${}^2E_2^{p,q} = {}''H^p({}'H^{q,p}(GI) = H^p(G^1H^{q,p}(I)).$$

By the full injectivity of the resolutions, the complex ${}'H^{q,p}(I)$ with $p \geqq 0$ is an injective resolution of

$$H^q(TC) = (R^qT)(A).$$

Furthermore, we have

$$H^p(G'H^{q,p}) = R^pG(R^qT(A)),$$

since a derived functor is the homology of an injective resolution. This proves that $(R^pG)R^qT(A))$ abuts to $R^n(GT)(A)$, and concludes the proof of the theorem.

Just to see the spectral sequence at work, we give one application relating it to the Euler characteristic discussed in §3.

Let $\mathfrak{A}$ have enough injectives, and let

$$T : \mathfrak{A} \to \mathfrak{B}$$

be a covariant left exact functor. Let $\mathfrak{F}_{\mathfrak{a}}$ be a family of objects in $\mathfrak{A}$ giving rise to a **K**-group. More precisely, in a short exact sequence in $\mathfrak{A}$, if two of the objects lie in $\mathfrak{F}_{\mathfrak{a}}$, then so does the third. We also assume that the objects of $\mathfrak{F}_{\mathfrak{a}}$ have **finite RT-dimension**, which means by definition that if $A \in \mathfrak{F}_{\mathfrak{a}}$ then $R^iT(A) = 0$

for all i sufficiently large. We could take $\mathfrak{F}_{\mathfrak{a}}$ in fact to be the family of all objects in $\mathfrak{A}$ which have finite RT-dimension.

We define the **Euler characteristic associated with** T on $\mathbf{K}(\mathfrak{F}_{\mathfrak{a}})$ to be

$$\chi_T(A) = \sum_{i=0}^{\infty} (-1)^i \operatorname{cl}(R^i T(A)).$$

The cl denotes the class in the **K**-group $K(\mathfrak{F}_{\mathfrak{B}})$ associated with some family $\mathfrak{F}_{\mathfrak{B}}$ of objects in $\mathfrak{B}$, and such that $R^iT(A) \in \mathfrak{F}_{\mathfrak{B}}$ for all $A \in \mathfrak{F}_{\mathfrak{a}}$. This is the minimum required for the formula to make sense.

Lemma 9.7. *The map χ_T extends to a homomorphism*

$$\mathbf{K}(\mathfrak{F}_{\mathfrak{a}}) \to \mathbf{K}(\mathfrak{F}_{\mathfrak{B}}).$$

Proof. Let

$$0 \to A' \to A \to A'' \to 0$$

be an exact sequence in $\mathfrak{F}$. Then we have the cohomology sequence

$$\to R^iT(A') \to R^iT(A) \to R^iT(A'') \to R^{i+1}T(A') \to$$

in which all but a finite number of terms are 0. Taking the alternating sum in the **K**-group shows that χ_T is an Euler–Poincaré map, and concludes the proof.

Note that we have merely repeated something from §3, in a jazzed up context. In the next theorem, we have another functor

$$G : \mathfrak{B} \to \mathfrak{C},$$

and we also have a family $\mathfrak{F}_{\mathfrak{C}}$ giving rise to a **K**-group $\mathbf{K}(\mathfrak{F}_{\mathfrak{C}})$. We suppose that we can perform the above procedure at each step, and also need some condition so that we can apply the spectral sequence. So, precisely, we assume:

CHAR 1. For all i, R^iT maps $\mathfrak{F}_{\mathfrak{a}}$ into $\mathfrak{F}_{\mathfrak{B}}$, R^iG maps $\mathfrak{F}_{\mathfrak{B}}$ into $\mathfrak{F}_{\mathfrak{C}}$, and $R^i(GT)$ maps $\mathfrak{F}_{\mathfrak{a}}$ into $\mathfrak{F}_{\mathfrak{C}}$.

CHAR 2. Each subobject of an element of $\mathfrak{F}_{\mathfrak{a}}$ lies in $\mathfrak{F}_{\mathfrak{a}}$ and has finite RT- and $R(GT)$-dimension; each subobject of an element of $\mathfrak{F}_{\mathfrak{B}}$ lies in $\mathfrak{F}_{\mathfrak{B}}$ and has finite RG-dimension.

Theorem 9.8. *Assume that $T : \mathfrak{A} \to \mathfrak{B}$ and $G : \mathfrak{B} \to \mathfrak{C}$ satisfy the conditions* **CHAR 1** *and* **CHAR 2**. *Also assume that T maps injectives to G-acyclics. Then*

$$\chi_G \circ \chi_T = \chi_{GT}.$$

Proof. By Theorem 9.6, the Grothendieck spectral sequence of the composite functor implies the existence of a filtration

$$\cdots \subset F^p R^n(GT)(A) \subset F^{p+1} R^n(GT)(A) \subset \cdots$$

of $R^n(GT)(A)$, such that

$$F^{p+1}/F^p \approx E_\infty^{p,n-p}.$$

Then

$$\begin{aligned}\chi_{GT}(A) &= \sum_{n=0}^{\infty} (-1)^n \operatorname{cl}(R^n(GT)(A)) \\ &= \sum_{n=0}^{\infty} (-1)^n \sum_{p=0}^{\infty} \operatorname{cl}(E_\infty^{p,n-p}) \\ &= \sum_{n=0}^{\infty} (-1)^n \operatorname{cl}(E_\infty^n).\end{aligned}$$

On the other hand,

$$\chi_T(A) = \sum_{q=0}^{\infty} (-1)^q \operatorname{cl}(R^q T(A))$$

and so

$$\begin{aligned}\chi_G \circ \chi_T(A) &= \sum_{q=0}^{\infty} (-1)^q \chi_G(R^q T(A)) \\ &= \sum_{q=0}^{\infty} (-1)^q \sum_{p=0}^{\infty} (-1)^p \operatorname{cl}(R^p G(R^q T(A)) \\ &= \sum_{n=0}^{\infty} (-1)^n \sum_{p=0}^{n} \operatorname{cl}(R^p G(R^{n-p} T(A)) \\ &= \sum_{n=0}^{\infty} (-1)^n \operatorname{cl}(E_2^n).\end{aligned}$$

Since E_{r+1} is the homology of E_r, we get

$$\sum_{n=0}^{\infty} (-1)^n \operatorname{cl}(E_2^n) = \sum_{n=0}^{\infty} (-1)^n \operatorname{cl}(E_3^n) = \cdots = \sum_{n=0}^{\infty} (-1)^n \operatorname{cl}(E_\infty^n).$$

This concludes the proof of the theorem.

EXERCISES

1. Prove that the example of the standard complex given in §1 is actually a complex, and is exact, so it gives a resolution of $\mathbf{Z}$. [*Hint:* To show that the sequence of the standard complex is exact, choose an element $z \in S$ and define $h : E^i \to E^{i+1}$ by letting

$$h(x_0, \ldots, x_i) = (z, x_0, \ldots, x_i).$$

Prove that $dh + hd = \text{id}$, and that $dd = 0$. Exactness follows at once.]

Cohomology of groups

2. Let G be a group. Use G as the set S in the standard complex. Define an action of G on the standard complex E by letting

$$x(x_0, \ldots, x_i) = (xx_0, \ldots, xx_i).$$

Prove that each E_i is a free module over the group ring $\mathbf{Z}[G]$. Thus if we let $R = \mathbf{Z}[G]$ be the group ring, and consider the category Mod(G) of G-modules, then the standard complex gives a free resolution of $\mathbf{Z}$ in this category.

3. The standard complex E was written in homogeneous form, so the boundary maps have a certain symmetry. There is another complex which exhibits useful features as follows. Let F^i be the free $\mathbf{Z}[G]$-module having for basis i-tuples (rather than $(i+1)$-tuples) $(x_1, \ldots, x_i)$. For $i = 0$ we take $F_0 = \mathbf{Z}[G]$ itself. Define the boundary operator by the formula

$$d(x_1, \ldots, x_i) = x_1(x_2, \ldots, x_i) + \sum_{j=1}^{i-1} (-1)^j (x_1, \ldots, x_j x_{j+1}, \ldots, x_i)$$
$$+ (-1)^{i+1}(x_1, \ldots, x_i).$$

Show that $E \approx F$ (as complexes of G-modules) via the association

$$(x_1, \ldots, x_i) \mapsto (1, x_1, x_1x_2, \ldots, x_1x_2 \cdots x_i),$$

and that the operator d given for F corresponds to the operator d given for E under this isomorphism.

4. If A is a G-module, let A^G be the submodule consisting of all elements $v \in A$ such that $xv = v$ for all $x \in G$. Thus A^G has trivial G-action. (This notation is convenient, but is *not* the same as for the induced module of Chapter XVIII.)
   (a) Show that if $H^q(G, A)$ denotes the q-th homology of the complex $\text{Hom}_G(E, A)$, then $H^0(G, A) = A^G$. Thus the left derived functors of $A \mapsto A^G$ are the homology groups of the complex $\text{Hom}_G(E, A)$, or for that matter, of the complex $\text{Hom}(F, A)$, where F is as in Exercise 3.
   (b) Show that the group of 1-cycles $Z^1(G, A)$ consists of those functions $f : G \to A$ satisfying

$$f(x) + xf(y) = f(xy) \text{ for all } x, y \in G.$$

Show that the subgroup of coboundaries $B^1(G, A)$ consists of those functions f for which there exists an element $a \in A$ such that $f(x) = xa - a$. The factor group is then $H^1(G, A)$. See Chapter VI, §10 for the determination of a special case.

(c) Show that the group of 2-cocycles $Z^2(G, A)$ consists of those functions $f: G \to A$ satisfying

$$xf(y, z) - f(xy, z) + f(x, yz) - f(x, y) = 0.$$

Such 2-cocycles are also called **factor sets**, and they can be used to describe isomorphism classes of group extensions, as follows.

5. **Group extensions**. Let W be a group and A a normal subgroup, written multiplicatively. Let $G = W/A$ be the factor group. Let F: $G \to W$ be a choice of coset representatives. Define

$$f(x, y) = F(x)F(y)F(xy)^{-1}.$$

(a) Prove that f is A-valued, and that $f: G \times G \to A$ is a 2-cocycle.
(b) Given a group G and an abelian group A, we view an extension W as an exact sequence

$$1 \to A \to W \to G \to 1.$$

Show that if two such extensions are isomorphic then the 2-cocycles associated to these extensions as in (a) define the same class in $H^1(G, A)$.
(c) Prove that the map which we obtained above from isomorphism classes of group extensions to $H^2(G, A)$ is a bijection.

6. **Morphisms of the cohomology functor.** Let $\lambda : G' \to G$ be a group homomorphism. Then λ gives rise to an exact functor

$$\Phi_\lambda : \mathrm{Mod}(G) \to \mathrm{Mod}(G'),$$

because every G-module can be viewed as a G'-module by defining the operation of $\sigma' \in G'$ to be $\sigma' a = \lambda(\sigma')a$. Thus we obtain a cohomology functor $H^{G'} \circ \Phi_\lambda$.

Let G' be a subgroup of G. In dimension 0, we have a morphism of functors

$$\lambda^* : H^0_G \to H^0_{G'} \circ \Phi_\lambda \text{ given by the inclusion } A^G \hookrightarrow A^{G'} = \Phi_\lambda(A)^{G'}.$$

(a) Show that there is a unique morphism of δ-functors

$$\lambda^* : H_G \to H_{G'} \circ \Phi_\lambda$$

which has the above effect on H^0_G. We have the following important special cases.

Restriction. Let H be a subgroup of G. Let A be a G-module. A function from G into A restricts to a function from H into A. In this way, we get a natural homomorphism called the **restriction**

$$\mathrm{res}: H^q(G, A) \to H^q(H, A).$$

Inflation. Suppose that H is normal in G. Let A^H be the subgroup of A consisting of those elements fixed by H. Then it is immediately verified that A^H is stable under G, and so is a G/H-module. The inclusion $A^H \hookrightarrow A$ induces a homomorphism

$$H^q_G(u) = u_q : H^q(G, A^H) \to H^q(A).$$

Define the **inflation**

$$\inf^H_{G/H} : H^q(G/H, A^H) \to H^q(G, A)$$

as the composite of the functorial morphism $H^q(G/H, A^H) \to H^q(G, A^H)$ followed by the induced homomorphism $u_q = H^q_G(u)$ as above.

In dimension 0, the inflation gives the identity $(A^H)^{G/H} = A^G$.

(b) Show that the inflation can be expressed on the standard cochain complex by the natural map which to a function of G/H in A^H associates a function of G into $A^H \subset A$.

(c) Prove that the following sequence is exact.

$$0 \to H^1(G/H, A^H) \xrightarrow{\text{inf}} H^1(G, A) \xrightarrow{\text{res}} H^1(H, A).$$

(d) Describe how one gets an operation of G on the cohomology functor H_G "by conjugation" and functoriality.

(e) In (c), show that the image of restriction on the right actually lies in $H^1(H, A)^G$ (the fixed subgroup under G).

Remark. There is an analogous result for higher cohomology groups, whose proof needs a spectral sequence of Hochschild-Serre. See [La 96], Chapter VI, §2, Theorem 2. It is actually this version for H^2 which is applied to $H^2(G, K^*)$, when K is a Galois extension, and is used in class field theory [ArT 67].

7. Let G be a group, B an abelian group and $M_G(B) = M(G, B)$ the set of mappings from G into B. For $x \in G$ and $f \in M(G, B)$ define $([x]f)(y) = f(yx)$.

(a) Show that $B \mapsto M_G(B)$ is a covariant, additive, exact functor from Mod($\mathbf{Z}$) (category of abelian groups) into Mod(G).

(b) Let G' be a subgroup of G and $G = \bigcup x_j G'$ a coset decomposition. For $f \in M(G, B)$ let f_j be the function in $M(G', B)$ such that $f_j(y) = f(x_j y)$. Show that the map

$$f \mapsto \prod_j f_j$$

is a G'-isomorphism from $M(G, B)$ to $\prod_j M(G', B)$.

8. For each G-module $A \in$ Mod(G), define $\varepsilon_A : A \to M(G, A)$ by the condition $\varepsilon_A(a) =$ the function f_a such that $f_z(\sigma) = \sigma a$ for $\sigma \in G$. Show that $a \mapsto f_a$ is a G-module embedding, and that the exact sequence

$$0 \to A \xrightarrow{\varepsilon_A} M(G, A) \to X_A = \operatorname{coker} \varepsilon_A \to 0$$

splits over $\mathbf{Z}$. (In fact, the map $f \mapsto f(e)$ splits the left side arrow.)

9. Let $B \in$ Mod($\mathbf{Z}$). Let H^q be the left derived functor of $A \mapsto A^G$.

(a) Show that $H^q(G, M_G(B)) = 0$ for all $q > 0$. [*Hint*: use a contracting homotopy

$$s: C^r(G, M_G(B)) \to C^{r-1}(G, M_G(B)) \quad \text{by} \quad (sf)_{x_2, \ldots, x_r}(x) = f_{x, x_2, \ldots, x_r}(1).$$

Show that $f = sdf + dsf$.] Thus M_G erases the cohomology functor.

(b) Also show that for all subgroups G' of G one has $H^q(G', M_G(B)) = 0$ for $q > 0$.

10. Let G be a group and S a subgroup. Show that the bifunctors

$$(A, B) \mapsto \operatorname{Hom}_G(A, M^S_G(B)) \text{ and } (A, B) \mapsto \operatorname{Hom}_S(A, B)$$

on Mod(G) $\times$ Mod(S) with value in Mod($\mathbf{Z}$) are isomorphic. The isomorphism is given by the maps

$$\varphi \mapsto (a \mapsto g_a), \text{ for } \varphi \in \operatorname{Hom}_S(A, B), \text{ where } g_a(\sigma) = \varphi(\sigma a),\ g_a \in M^S_G(B).$$

The inverse mapping is given by

$$f \mapsto f(1) \text{ with } f \in \mathrm{Hom}_G(A, M_G^S(B)).$$

Recall that $M_G^S(B)$ was defined in Chapter XVIII, §7 for the induced representation. Basically you should already know the above isomorphism.

11. Let G be a group and S a subgroup. Show that the map

$$H^q(G, M_G^S(B)) \to H^q(S, B) \text{ for } B \in \mathrm{Mod}(S),$$

obtained by composing the restriction res_S^G with the S-homomorphism $f \mapsto f(1)$, is an isomorphism for $q > 0$. [*Hint*: Use the uniqueness theorem for cohomology functors.]

12. Let G be a group. Let $\varepsilon : \mathbf{Z}[G] \to \mathbf{Z}$ be the homomorphism such that $\varepsilon\left(\sum n(x)x\right) = \sum n(x)$. Let I_G be its kernel. Prove that I_G is an ideal of $\mathbf{Z}[G]$ and that there is an isomorphism of functors (on the category of groups)

$$G/G^c \approx I_G/I_G^2, \qquad \text{by} \qquad xG^c \mapsto (x - 1) + I_G^2.$$

13. Let $A \in \mathrm{Mod}(G)$ and $\alpha \in H^1(G, A)$. Let $\{a(x)\}_{x \in G}$ be a standard 1-cocycle representing α. Show that there exists a G-homomorphism $f : I_G \to A$ such that $f(x - 1) = a(x)$, so $f \in (\mathrm{Hom}(I_G, A))^G$. Show that the sequence

$$0 \to A = \mathrm{Hom}(\mathbf{Z}, A) \to \mathrm{Hom}(\mathbf{Z}[G], A) \to \mathrm{Hom}(I_G, A) \to 0$$

is exact, and that if δ is the coboundary for the cohomology sequence, then $\delta(f) = -\alpha$.

Finite groups

We now turn to the case of *finite* groups G. For such groups and a G-module A we have the **trace**

$$T_G : A \to A \qquad \text{defined by} \qquad T_G(a) = \sum_{\sigma \in G} \sigma a.$$

We define a module A to be G-**regular** if there exists a $\mathbf{Z}$-endomorphism $u : A \to A$ such that $\mathrm{id}_A = T_G(u)$. Recall that the operation of G on $\mathrm{End}(A)$ is given by

$$[\sigma]f(a) = \sigma f(\sigma^{-1}a) \text{ for } \sigma \in G.$$

14. (a) Show that a projective object in $\mathrm{Mod}(G)$ is G-regular.
    (b) Let R be a commutative ring and let A be in $\mathrm{Mod}_R(G)$ (the category of (G, R)-modules). Show that A is $R[G]$-projective if and only if A is R-projective and $R[G]$-regular, meaning that $\mathrm{id}_A = T_G(u)$ for some R-homomorphism $u : A \to A$.

15. Consider the exact sequences:

(1) $$0 \to I_G \to \mathbf{Z}[G] \xrightarrow{\varepsilon} \mathbf{Z} \to 0$$

(2) $$0 \to \mathbf{Z} \xrightarrow{\varepsilon'} \mathbf{Z}[G] \to J_G \to 0$$

where the first one defines I_G, and the second is defined by the embedding

$$\varepsilon' : \mathbf{Z} \to \mathbf{Z}[G] \text{ such that } \varepsilon'(n) = n\left(\sum \sigma\right),$$

i.e. on the "diagonal". The cokernel of ε' is J_G by definition.

(a) Prove that both sequences (1) and (2) split in $\mathrm{Mod}(G)$.

(b) Define $M'_G(A) = \mathbf{Z}[G] \otimes A$ (tensor product over $\mathbf{Z}$) for $A \in \text{Mod}(G)$. Show that $M'_G(A)$ is G-regular, and that one gets exact sequences (1_A) and (2_A) by tensoring (1) and (2) with A. As a result one gets an embedding

$$\varepsilon'_A = \varepsilon' \otimes \text{id} : A = \mathbf{Z} \otimes A \to \mathbf{Z}[G] \otimes A.$$

16. **Cyclic groups**. Let G be a finite cyclic group of order n. Let σ be a generator of G. Let $K^i = \mathbf{Z}[G]$ for $i > 0$. Let $\varepsilon : K^0 \to \mathbf{Z}$ be the augmentation as before. For i odd $\geqq 1$, let $d^i : K^i \to K^{i-1}$ be multiplication by $1 - \sigma$. For i even $\geqq 2$, let d^i be multiplication by $1 + \sigma + \cdots + \sigma^{n-1}$. Prove that K is a resolution of $\mathbf{Z}$. Conclude that:
    For i odd: $H^i(G, A) = A^G/T_G A$ where $T_G : a \mapsto (1 + \sigma + \cdots + \sigma^{n-1})a$;
    For i even $\geqq 2$: $H^i(G, A) = A_T/(1 - \sigma)A$, where A_T is the kernel of T_G in A.

17. Let G be a finite group. Show that there exists a δ-functor $\mathbf{H}$ from $\text{Mod}(G)$ to $\text{Mod}(\mathbf{Z})$ such that:
    (1) $\mathbf{H}^0$ is (isomorphic to) the functor $A \mapsto A^G/T_G A$.
    (2) $\mathbf{H}^q(A) = 0$ if A is injective and $q > 0$, and $\mathbf{H}^q(A) = 0$ if A is projective and q is arbitrary.
    (3) $\mathbf{H}$ is erased by G-regular modules. In particular, $\mathbf{H}$ is erased by M_G.
    The δ-functor of Exercise 17 is called the **special cohomology functor**. It differs from the other one only in dimension 0.

18. Let $\mathbf{H} = \mathbf{H}_G$ be the special cohomology functor for a finite group G. Show that:

$$\mathbf{H}^0(I_G) = 0;\ \mathbf{H}^0(Z) \approx \mathbf{H}^1(I) \approx \mathbf{Z}/n\mathbf{Z} \text{ where } n = \#(G);$$

$$\mathbf{H}^0(Q/Z) = \mathbf{H}^1(Z) = \mathbf{H}^2(I) = 0$$

$$\mathbf{H}^1(Q/Z) \approx \mathbf{H}^2(Z) \approx \mathbf{H}^3(I) \approx G^\wedge = \text{Hom}(G, \mathbf{Q}/\mathbf{Z}) \text{ by definition.}$$

Injectives

19. (a) Show that if an abelian group T is injective in the category of abelian groups, then it is divisible.
    (b) Let A be a principal entire ring. Define the notion of divisibility by elements of A for modules in a manner analogous to that for abelian groups. Show that an A-module is injective if and only if it is A-divisible. [The proof for $\mathbf{Z}$ should work in exactly the same way.]

20. Let S be a multiplicative subset of the commutative Noetherian ring A. If I is an injective A-module, show that $S^{-1}I$ is an injective $S^{-1}A$-module.

21. (a) Show that a direct sum of projective modules is projective.
    (b) Show that a direct product of injective modules is injective.

22. Show that a factor module, direct summand, direct product, and direct sum of divisible modules are divisible.

23. Let Q be a module over a commutative ring A. Assume that for every left ideal J of A, every homomorphism $\varphi : J \to Q$ can be extended to a homomorphism of A into Q. Show that Q is injective. [*Hint*: Given $M' \subset M$ and $f : M' \to Q$, let $x_0 \in M$ and $x_0 \notin M'$. Let J be the left ideal of elements $a \in A$ such that $ax_0 \in M'$. Let $\varphi(a) = f(ax_0)$ and extend φ to A, as can be done by hypothesis. Then show that

one can extend f to M by the formula

$$f(x' + bx_0) = f(x') + \varphi(b),$$

for $x' \in M$ and $b \in A$. Then use Zorn's lemma. This is the same pattern of proof as the proof of Lemma 4.2.]

24. Let

$$0 \to I_1 \to I_2 \to I_3 \to 0$$

be an exact sequence of modules. Assume that I_1, I_2 are injective.
   (a) Show that the sequence splits.
   (b) Show that I_3 is injective.
   (c) If I is injective and $I = M \oplus N$, show that M is injective.

25. (Do this exercise after you have read about Noetherian rings.) Let A be a Noetherian commutative ring, and let Q be an injective A-module. Let $\mathfrak{a}$ be an ideal of A, and let $Q^{(\mathfrak{a})}$ be the subset of elements $x \in Q$ such that $\mathfrak{a}^n x = 0$ for some n, depending on x. Show that $Q^{(\mathfrak{a})}$ is injective. [*Hint*: Use Exercise 23.]

26. Let A be a commutative ring. Let E be an A-module, and let $E^\wedge = \mathrm{Hom}_{\mathbf{Z}}(E, \mathbf{Q}/\mathbf{Z})$ be the dual module. Prove the following statements.
   (a) A sequence

$$0 \to N \to M \to E \to 0$$

   is exact if and only if the dual sequence

$$0 \to E^\wedge \to M^\wedge \to N^\wedge \to 0$$

   is exact.
   (b) Let F be flat and I injective in the category of A-modules. Show that $\mathrm{Hom}_A(F, I)$ is injective.
   (c) E is flat if and only if $E^\wedge$ is injective.

27. **Extensions of modules**. Let M, N be modules over a ring. By an **extension** of M by N we mean an exact sequence

$$(*) \qquad 0 \to N \to E \to M \to 0.$$

We shall now define a map from such extensions to $\mathrm{Ext}^1(M, N)$. Let P be projective, with a surjective homomorphism onto M, so we get an exact sequence

$$(**) \qquad 0 \to K \xrightarrow{w} P \xrightarrow{p} M \to 0$$

where K is defined to be the kernel. Since P is projective, there exists a homomorphism $u: P \to E$, and depending on u a unique homomorphism $v: K \to N$ making the diagram commutative:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & K & \longrightarrow & P & \longrightarrow & M & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle v} & & \downarrow{\scriptstyle u} & & \downarrow{\scriptstyle \mathrm{id}} & & \\ 0 & \longrightarrow & N & \longrightarrow & E & \longrightarrow & M & \longrightarrow & 0 \end{array}$$

On the other hand, we have the exact sequence

(***) $\quad 0 \to \mathrm{Hom}(M, N) \to \mathrm{Hom}(P, N) \to \mathrm{Hom}(K, N) \to \mathrm{Ext}^1(M, N) \to 0,$

with the last term on the right being equal to 0 because $\mathrm{Ext}^1(P, N) = 0$. To the extension (*) we associate the image of v in $\mathrm{Ext}^1(M, N)$.

Prove that this association is a bijection between isomorphism classes of extensions (i.e. isomorphism classes of exact sequences as in (*)), and $\mathrm{Ext}^1(M, N)$. [*Hint*: Construct an inverse as follows. Given an element e of $\mathrm{Ext}^1(M, N)$, using an exact sequence (**), there is some element $v \in \mathrm{Hom}(K, N)$ which maps on e in (***). Let E be the push-out of v and w. In other words, let J be the submodule of $N \oplus P$ consisting of all elements $(v(x), -w(x))$ with $x \in K$, and let $E = (N \oplus P)/J$. Show that the map $y \mapsto (y, 0) \bmod J$ gives an injection of N into E. Show that the map $N \oplus P \to M$ vanishes on J, and so gives a surjective homomorphism $E \to M \to 0$. Thus we obtain an exact sequence (*); that is, an extension of M by N. Thus to each element of $\mathrm{Ext}^1(M, N)$ we have associated an isomorphism class of extensions of M by N. Show that the maps we have defined are inverse to each other between isomorphism classes of extensions and elements of $\mathrm{Ext}^1(M, N)$.]

28. Let R be a principal entire ring. Let $a \in R$. For every R-module N, prove:
    (a) $\mathrm{Ext}^1(R/aR, N) = N/aN$.
    (b) For $b \in R$ we have $\mathrm{Ext}^1(R/aR, R/bR) = R/(a, b)$, where (a, b) is the g.c.d of a and b, assuming $ab \neq 0$.

Tensor product of complexes.

29. Let $K = \bigoplus K_p$ and $L = \bigoplus L_q$ be two complexes indexed by the integers, and with boundary maps lower indices by 1. Define $K \otimes L$ to be the direct sum of the modules $(K \otimes L)_n$, where

$$(K \otimes L)_n = \bigoplus_{p+q=n} K_p \otimes L_q.$$

Show that there exist unique homomorphisms

$$d = d_n : (K \otimes L)_n \to (K \otimes L)_{n-1}$$

such that

$$d(x \otimes y) = d(x) \otimes y + (-1)^p x \otimes d(y).$$

Show that $K \otimes L$ with these homomorphisms is a complex, that is $d \circ d = 0$.

30. Let K, L be double complexes. We write K_i and L_i for the ordinary column complexes of K and L respectively. Let $\varphi: K \to L$ be a homomorphism of double complexes. Assume that each homomorphism

$$\varphi_i: K_i \to L_i$$

is a homology isomorphism.
    (a) Prove that $\mathrm{Tot}(\varphi): \mathrm{Tot}(K) \to \mathrm{Tot}(L)$ is a homology isomorphism. (If you want to see this worked out, cf. [FuL 85], Chapter V, Lemma 5.4.)
    (b) Prove Theorem 9.8 using (a) instead of spectral sequences.

Bibliography

[ArT 68] E. ARTIN and J. TATE, *Class Field Theory*, Benjamin, 1968; Addison-Wesley, 1991

[At 61] M. ATIYAH, Characters and cohomology of finite groups, *Pub. IHES* **9** (1961), pp. 5–26

[At 67] M. ATIYAH, *K-theory*, Benjamin, 1967; reprinted Addison-Wesley, 1991

[ABP 73] M. ATIYAH, R. BOTT, and R. PATODI, On the heat equation and the index theorem, *Invent. Math.* **19** (1973), pp. 279–330

[Ba 68] H. BASS, *Algebraic K-theory*, Benjamin, 1968

[Bo 69] R. BOTT, *Lectures on K(X)*, Benjamin, 1969

[BtD 85] T. BROCKER and T. TOM DIECK, *Representations of Compact Lie Groups*, Springer Verlag, 1985

[CaE 57] H. CARTAN and S. EILENBERG, *Homological Algebra*, Princeton University Press, 1957

[CuR 81] C. CURTIS and I. REINER, *Methods of Representation Theory*, John Wiley & Sons, 1981

[ES 52] S. EILENBERG and N. STEENROD, *Foundations of Algebraic Topology*, Princeton University Press, 1952

[FuL 85] W. FULTON and S. LANG, *Riemann-Roch algebra*, Springer Verlag, 1985

[Go 58] R. GODEMENT, *Théorie des faisceaux*, Hermann Paris, 1958

[GreH 81] M. GREENBERG and J. HARPER, *Algebraic Topology: A First Course*, Benjamin-Addison-Wesley, 1981

[GriH 78] P. GRIFFITHS and J. HARRIS, *Principles of algebraic geometry*, Wiley Interscience 1978

[Gro 57] A. GROTHENDIECK, Sur quelques points d'algèbre homologique, *Tohoku Math. J.* **9** (1957) pp. 119–221

[Gro 68] A. GROTHENDIECK, *Classes de Chern et représentations linéaires des groupes discrets*, Dix exposés sur la cohomologie étale des schémas, North-Holland, Amsterdam, 1968

[Gu 91] R. GUNNING, *Introduction to holomorphic functions of several variables*, Vol. III Wadsworth & Brooks/Cole, 1990

[Ha 77] R. HARTSHORNE, *Algebraic Geometry*, Springer Verlag, 1977

[HiS 70] P. J. HILTON and U. STAMMBACH, *A Course in Homological Algebra*, Graduate Texts in Mathematics, Springer Verlag, 1970.

[La 96] S. LANG, *Topics in cohomology of groups*, Springer Lecture Notes, 1996

[Man 69] J. MANIN, *Lectures on the K-functor in Algebraic Geometry, Russian Math Surveys* **24**(5) (1969) pp. 1–89

[Mat 70] H. MATSUMURA, *Commutative Algebra*, Second Edition, Benjamin-Cummings, 1981

[No 68] D. NORTHCOTT, *Lessons on Rings, Modules and Multiplicities*, Cambridge University Press, 1968

[No 76] D. NORTHCOTT, *Finite Free Resolutions*, Cambridge University Press, 1976

[Ro 79] J. ROTMAN, *Introduction to Homological Algebra*, Academic Press, 1979

[Se 64] J.-P. SERRE, *Cohomologie Galoisienne*, Springer Lecture Notes **5**, 1964

[Se 65] J.-P. SERRE, *Algèbre locale, multiplicités*, Springer Lecture Notes **11** (1965) Third Edition 1975

[SGA 6] P. BERTHELOT, A. GROTHENDIECK, L. ILLUSIE et al. *Théorie des intersections et théorème de Riemann-Roch*, Springer Lecture Notes 146, 1970

[Sh 72] S. SHATZ, *Profinite groups, arithmetic and geometry*, Ann. of Math Studies, Princeton University Press 1972

CHAPTER XXI

Finite Free Resolutions

This chapter puts together specific computations of complexes and homology. Partly these provide examples for the general theory of Chapter XX, and partly they provide concrete results which have occupied algebraists for a century. They have one aspect in common: the computation of homology is done by means of a finite free resolution, i.e. a finite complex whose modules are finite free.

The first section shows a general technique (the mapping cylinder) whereby the homology arising from some complex can be computed by using another complex which is finite free. One application of such complexes has already been given in Chapter X, putting together Proposition 4.5 followed by Exercises 10–15 of that chapter.

Then we go to major theorems, going from Hilbert's Syzygy theorem, from a century ago, to Serre's theorem about finite free resolutions of modules over polynomial rings, and the Quillen-Suslin theorem. We also include a discussion of certain finite free resolutions obtained from the Koszul complex. These apply, among other things, to the Grothendieck Riemann-Roch theorem of algebraic geometry.

Bibliographical references refer to the list given at the end of Chapter XX.

§1. SPECIAL COMPLEXES

As in the preceding chapter, we work with the category of modules over a ring, but the reader will notice that the arguments hold quite generally in an abelian category.

In some applications one determines homology from a complex which is not suitable for other types of construction, like changing the base ring. In this section, we give a general procedure which constructs another complex with

better properties than the first one, while giving the same homology. For an application to Noetherian modules, see Exercises 12–15 of Chapter X.

Let $f: K \to C$ be a morphism of complexes. We say that f is a **homology isomorphism** if the natural map

$$H(f): H(K) \to H(C)$$

is an isomorphism. The definition is valid in an abelian category, but the reader may think of modules over a ring, or abelian groups even. A family $\mathfrak{F}$ of objects will be called **sufficient** if given an object E there exists an element F in $\mathfrak{F}$ and an epimorphism

$$F \to E \to 0,$$

and if $\mathfrak{F}$ is closed under taking finite direct sums. For instance, we may use for $\mathfrak{F}$ the family of free modules. However, in important applications, we shall deal with finitely generated modules, in which case $\mathfrak{F}$ might be taken as the family of finite free modules. These are in fact the applications I have in mind, which resulted in having axiomatized the situation.

Proposition 1.1. *Let C be a complex such that $H^p(C) \neq 0$ only for $0 \leqq p \leqq n$. Let $\mathfrak{F}$ be a sufficient family of projectives. There exists a complex*

$$0 \to K^0 \to K^1 \to \cdots \to K^n \to 0$$

such that:

$$K^p \neq 0 \quad \text{only for} \quad 0 \leqq p \leqq n;$$

$$K^p \text{ is in } \mathfrak{F} \text{ for all } p \geqq 1;$$

and there exists a homomorphism of complexes

$$f: K \to C$$

which is a homology isomorphism.

Proof. We define f_m by descending induction on m:

$$\begin{array}{ccccccccc}
\longrightarrow & K^m & \longrightarrow & K^{m+1} & \xrightarrow{\delta_K^{m+1}} & K^{m+2} & \longrightarrow \\
 & \downarrow{\scriptstyle f_m} & & \downarrow{\scriptstyle f_{m+1}} & & \downarrow{\scriptstyle f_{m+2}} & \\
\longrightarrow & C^m & \longrightarrow & C^{m+1} & \xrightarrow[\delta_C^{m+1}]{} & C^{m+2} & \longrightarrow
\end{array}$$

We suppose that we have defined a morphism of complexes with $p \geqq m + 1$ such that $H^p(f)$ is an isomorphism for $p \geqq m + 2$, and

$$f_{m+1}: Z^{m+1}(K) \to H^{m+1}(C)$$

is an epimorphism, where Z denotes the cycles, that is Ker δ. We wish to construct K^m and f_m, thus propagating to the left. First let $m \geqq 0$. Let B^{m+1} be the kernel of

$$\text{Ker } \delta_K^{m+1} \to H^{m+1}(C).$$

Let K' be in $\mathfrak{F}$ with an epimorphism

$$\delta' : K' \to B^{m+1}.$$

Let $K'' \to H^m(C)$ be an epimorphism with K'' in $\mathfrak{F}$, and let

$$f'' : K'' \to Z^m(C)$$

be any lifting, which exists since K'' is projective. Let

$$K^m = K' \oplus K''$$

and define $\delta^m : K^m \to K^{m+1}$ to be δ' on K' and 0 on K''. Then

$$f_{m+1} \circ \delta'(K') \subset \delta_C(C_m),$$

and hence there exists $f' : K' \to C^m$ such that

$$\delta_C \circ f' = f_{m+1} \circ \delta'.$$

We now define $f_m : K^m \to C^m$ to be f' on K' and f'' on K''. Then we have defined a morphism of complexes truncated down to m as desired.

Finally, if $m = -1$, we have constructed down to K^0, δ^0, and f_0 with

$$K^0 \xrightarrow{f_0} H^0(C) \to 0$$

exact. The last square looks like this, defining $K^{-1} = 0$.

$$\begin{array}{ccccc} 0 & \longrightarrow & K' \oplus K'' & \xrightarrow{\delta_0 = \delta'} & \delta' K' \subset K^1 \\ & & f' \searrow \swarrow f'' & & \downarrow f_1 \\ 0 & \longrightarrow & C^0 & \longrightarrow & C^1 \end{array}$$

We replace K^0 by $K^0/(\text{Ker } \delta^0 \cap \text{Ker} f_0)$. Then $H^0(f)$ becomes an isomorphism, thus proving the proposition.

We want to say something more about K^0. For this purpose, we define a new concept. Let $\mathfrak{F}$ be a family of objects in the given abelian category (think of modules in first reading). We shall say that $\mathfrak{F}$ is **complete** if it is sufficient, and for any exact sequence

$$0 \to F' \to F \to F'' \to 0$$

with F'' and F in $\mathfrak{F}$ then F' is also in $\mathfrak{F}$.

Example. In Chapter XVI, Theorem 3.4 we proved that the family of finite flat modules in the category of finite modules over a Noetherian ring is complete. Similarly, the family of flat modules in the category of modules over a ring is complete. We cannot get away with just projectives or free modules, because in the statement of the proposition, K^0 is not necessarily free but we want to include it in the family as having especially nice properties. In practice, the family consists of the flat modules, or finite flat modules. Cf. Chaper X, Theorem 4.4, and Chapter XVI, Theorem 3.8.

Proposition 1.2. *Let $f: K \to C$ be a morphism of complexes, such that K^p, $H^p(C)$ are $\neq 0$ only for $p = 1, \dots, n$. Let $\mathfrak{F}$ be a complete family, and assume that K^p, C^p are in $\mathfrak{F}$ for all p, except possibly for K^0. If f is a homology isomorphism, then K^0 is also in $\mathfrak{F}$.*

Before giving the proof, we define a new complex called the **mapping cylinder** of an arbitrary morphism of complexes f by letting

$$M^p = K^p \oplus C^{p-1}$$

and defining $\delta_M : M^p \to M^{p+1}$ by

$$\delta_M(x, y) = (\delta x, fx - \delta y).$$

It is trivially verified that M is then a complex, i.e. $\delta \circ \delta = 0$. If C' is the complex obtained from C by shifting degrees by one (and making a sign change in δ_C), so $C'^p = C^{p-1}$, then we get an exact sequence of complexes

$$0 \to C' \to M \to K \to 0$$

and hence the **mapping cylinder exact cohomology sequence**

$$\begin{array}{ccccccccc} H^p(K) & \longrightarrow & H^{p+1}(C') & \longrightarrow & H^{p+1}(M) & \longrightarrow & H^{p+1}(K) & \longrightarrow & H^{p+2}(C') \\ & & \| & & & & & & \| \\ & & H^p(C) & & & & & & H^{p+1}(C) \end{array}$$

and one sees from the definitions that the cohomology maps

$$H^p(K) \to H^{p+1}(C') \approx H^p(C)$$

are the ones induced by $f: K \to C$.

We now return to the assumptions of Proposition 1.2, so that these maps are isomorphisms. We conclude that $H(M) = 0$. This implies that the sequence

$$0 \to K^0 \to M^1 \to M^2 \to \cdots \to M^{n+1} \to 0$$

is exact. Now each M^p is in $\mathfrak{F}$ by assumption. Inserting the kernels and cokernels at each step and using induction together with the definition of a complete family, we conclude that K^0 is in $\mathfrak{F}$, as was to be shown.

In the next proposition, we have axiomatized the situation so that it is applicable to the tensor product, discussed later, and to the case when the family $\mathfrak{F}$ consists of flat modules, as defined in Chapter XVI. No knowledge of this chapter is needed here, however, since the axiomatization uses just the general language of functors and exactness.

Let $\mathfrak{F}$ be a complete family again, and let T be a covariant additive functor on the given category. We say that $\mathfrak{F}$ is **exact for** T if given an exact sequence

$$0 \to F' \to F \to F'' \to 0$$

in $\mathfrak{F}$, then

$$0 \to T(F') \to T(F) \to T(F'') \to 0$$

is exact.

Proposition 1.3. *Let $\mathfrak{F}$ be a complete family which is exact for T. Let $f: K \to C$ be a morphism of complexes, such that K^p and C^p are in $\mathfrak{F}$ for all p, and K^p, $H^p(C)$ are zero for all but a finite number of p. Assume that f is a homology isomorphism. Then*

$$T(f): T(K) \to T(C)$$

is a homology isomorphism.

Proof. Construct the mapping cylinder M for f. As in the proof of Proposition 1.2, we get $H(M) = 0$ so M is exact. We then start inductively from the right with zeros. We let Z^p be the cycles in M^p and use the short exact sequences

$$0 \to Z^p \to M^p \to Z^{p+1} \to 0$$

together with the definition of a complete family to conclude that Z^p is in $\mathfrak{F}$ for all p. Hence the short sequences obtained by applying T are exact. But $T(M)$ is the mapping cylinder of the morphism

$$T(f): T(K) \to T(C),$$

which is therefore an isomorphism, as one sees from the homology sequence of the mapping cylinder. This concludes the proof.

§2. FINITE FREE RESOLUTIONS

The first part of this section develops the notion of resolutions for a case somewhat more subtle than projective resolutions, and gives a good example for the considerations of Chapter XX. Northcott in [No 76] pointed out that minor adjustments of standard proofs also applied to the non-Noetherian rings, only occasionally slightly less tractable than the Noetherian ones.

Let A be a ring. A module E is called **stably free** if there exists a finite free module F such that $E \oplus F$ is finite free, and thus isomorphic to $A^{(n)}$ for some positive integer n. In particular, E is projective and finitely generated.

We say that a module M has a **finite free resolution** if there exists a resolution

$$0 \to E_n \to \cdots \to E_0 \to M \to 0$$

such that each E_i is finite free.

Theorem 2.1. *Let M be a projective module. Then M is stably free if and only if M admits a finite free resolution.*

Proof. If M is stably free then it is trivial that M has a finite free resolution. Conversely assume the existence of the resolution with the above notation. We prove that M is stably free by induction on n. The assertion is obvious if $n = 0$. Assume $n \geqq 1$. Insert the kernels and cokernels at each step, in the manner of dimension shifting. Say

$$M_1 = \operatorname{Ker}(E_0 \to P),$$

giving rise to the exact sequence

$$0 \to M_1 \to E_0 \to M \to 0.$$

Since M is projective, this sequence splits, and $E_0 \approx M \oplus M_1$. But M_1 has a finite free resolution of length smaller than the resolution of M, so there exists a finite free module F such that $M_1 \oplus F$ is free. Since $E_0 \oplus F$ is also free, this concludes the proof of the theorem.

A resolution

$$0 \to E_n \to \cdots \to E_0 \to M \to 0$$

is called **stably free** if all the modules E_i $(i = 0, \ldots, n)$ are stably free.

Proposition 2.2. *Let M be an A-module. Then M has a finite free resolution of length $n \geqq 1$ if and only if M has a stably free resolution of length n.*

Proof. One direction is trivial, so we suppose given a stably free resolution with the above notation. Let $0 \leqq i < n$ be some integer, and let F_i, F_{i+1} be finite free such that $E_i \oplus F_i$ and $E_{i+1} \oplus F_{i+1}$ are free. Let $F = F_i \oplus F_{i+1}$. Then we can form an exact sequence

$$0 \to E_n \to \cdots \to E_{i+1} \oplus F \to E_i \oplus F \to \cdots \to E_0 \to M \to 0$$

in the obvious manner. In this way, we have changed two consecutive modules in the resolution to make them free. Proceeding by induction, we can then make E_0, E_1 free, then E_1, E_2 free, and so on to conclude the proof of the proposition.

The next lemma is designed to facilitate dimension shifting.

We say that two modules M_1, M_2 are **stably isomorphic** if there exist finite free modules F_1, F_2 such that $M_1 \oplus F_1 \approx M_2 \oplus F_2$.

Lemma 2.3. *Let M_1 be stably isomorphic to M_2. Let*

$$0 \to N_1 \to E_1 \to M_1 \to 0$$

$$0 \to N_2 \to E_2 \to M_2 \to 0$$

be exact sequences, where M_1 is stably isomorphic to M_2, and E_1, E_2 are stably free. Then N_1 is stably isomorphic to N_2.

Proof. By definition, there is an isomorphism $M_1 \oplus F_1 \approx M_2 \oplus F_2$. We have exact sequences

$$0 \to N_1 \to E_1 \oplus F_1 \to M_1 \oplus F_1 \to 0$$

$$0 \to N_2 \to E_2 \oplus F_2 \to M_2 \oplus F_2 \to 0$$

By Schanuel's lemma (see below) we conclude that

$$N_1 \oplus E_2 \oplus F_2 \approx N_2 \oplus E_1 \oplus F_1.$$

Since E_1, E_2, F_1, F_2 are stably free, we can add finite free modules to each side so that the summands of N_1 and N_2 become free, and by adding 1-dimensional free modules if necessary, we can preserve the isomorphism, which proves that N_1 is stably isomorphic to N_2.

We still have to take care of **Schanuel's lemma**:

Lemma 2.4. *Let*

$$0 \to K \to P \to M \to 0$$

$$0 \to K' \to P' \to M \to 0$$

be exact sequences where P, P' are projective. Then there is an isomorphism

$$K \oplus P' \approx K' \oplus P.$$

Proof. Since P is projective, there exists a homomorphism $P \to P'$ making the right square in the following diagram commute.

$$\begin{array}{ccccccccc} 0 & \longrightarrow & K & \xrightarrow{i} & P & \longrightarrow & M & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle u} & & \downarrow{\scriptstyle w} & & \downarrow{\scriptstyle \mathrm{id}} & & \\ 0 & \longrightarrow & K' & \xrightarrow[j]{} & P' & \longrightarrow & M & \longrightarrow & 0 \end{array}$$

Then one can find a homomorphism $K \to K'$ which makes the left square commute. Then we get an exact sequence

$$0 \to K \to P \oplus K' \to P' \to 0$$

by $x \mapsto (ix, ux)$ for $x \in K$ and $(y, z) \mapsto wy - jz$. We leave the verification of exactness to the reader. Since P' is projective, the sequence splits thus proving Schanuel's lemma. This also concludes the proof of Lemma 2.3.

The minimal length of a stably free resolution of a module is called its **stably free dimension**. To construct a stably free resolution of a finite module, we proceed inductively. The preceding lemmas allow us to carry out the induction, and also to stop the construction if a module is of finite stably free dimension.

Theorem 2.5. *Let M be a module which admits a stably free resolution of length n*

$$0 \to E_n \to \cdots \to E_0 \to M \to 0.$$

Let

$$F_m \to \cdots \to F_0 \to M \to 0$$

be an exact sequence with F_i stably free for $i = 0, \ldots, m$.

(i) *If $m < n - 1$ then there exists a stably free F_{m+1} such that the exact sequence can be continued exactly to*

$$F_{m+1} \to \cdots \to F_0 \to M \to 0.$$

(ii) *If $m = n - 1$, let $F_n = \mathrm{Ker}(F_{n-1} \to F_{n-2})$. Then F_n is stably free and thus*

$$0 \to F_n \to F_{n-1} \to \cdots \to F_0 \to M \to 0$$

is a stably free resolution.

Remark. If A is Noetherian then of course (i) is trivial, and we can even pick F_{m+1} to be finite free.

Proof. Insert the kernels and cokernels in each sequence, say

$$K_m = \mathrm{Ker}(E_m \to E_{m-1}) \quad \text{if} \quad m \neq 0$$

$$K_0 = \mathrm{Ker}(E_0 \to M),$$

and define K'_m similarly. By Lemma 2.3, K_m is stably isomorphic to K'_m, say

$$K_m \oplus F \approx K'_m \oplus F'$$

with F, F' finite free.

If $m < n - 1$, then K_m is a homomorphic image of E_{m+1}; so both $K_m \oplus F$ and $K'_m \oplus F'$ are homomorphic images of $E_{m+1} \oplus F$. Therefore K'_m is a homomorphic image of $E_{m+1} \oplus F$ which is stably free. We let $F_{m+1} = E_{m+1} \oplus F$ to conclude the proof in this case.

If $m = n - 1$, then we can take $K_n = E_n$. Hence $K_m \oplus F$ is stably free, and so is $K'_m \oplus F'$ by the isomorphism in the first part of the proof. It follows trivially that K'_m is stably free, and by definition, $K'_m = F_{m+1}$ in this case. This concludes the proof of the theorem.

Corollary 2.6. *If $0 \to M_1 \to E \to M \to 0$ is exact, M has stably free dimension $\leqq n$, and E is stably free, then M_1 has stably free dimension $\leqq n - 1$.*

Theorem 2.7. *Let*

$$0 \to M' \to M \to M'' \to 0$$

be an exact sequence. If any two of these modules have a finite free resolution, then so does the third.

Proof. Assume M' and M have finite free resolutions. Since M is finite, it follows that M'' is also finite. By essentially the same construction as Chapter XX, Lemma 3.8, we can construct an exact and commutative diagram where E', E, E'' are stably free:

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & 0 & & \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & M'_1 & \to & M_1 & \to & M''_1 & \to & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & E' & \to & E & \to & E'' & \to & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & M' & \to & M & \to & M'' & \to & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & & \\
 & & 0 & & 0 & & 0 & &
\end{array}$$

We then argue by induction on the stably free dimension of M. We see that M_1 has stably free dimension $\leqq n - 1$ (actually $n - 1$, but we don't care), and M'_1 has finite stably free dimension. By induction we are reduced to the case when M has stably free dimension 0, which means that M is stably free. Since by assumption there is a finite free resolution of M', it follows that M'' also has a finite free resolution, thus concluding the proof of the first assertion.

Next assume that M', M'' have finite free resolutions. Then M is finite. If both M' and M'' have stably free dimension 0, then M', M'' are projective and $M \approx M' \oplus M''$ is also stably free and we are done. We now argue by induction on the maximum of their stably free dimension n, and we assume $n \geqq 1$. We can construct an exact and commutative diagram as in the previous case with E', E, E'' finite free (we leave the details to the reader). But the maximum of the stably free dimensions of M'_1 and M''_1 is at most $n - 1$, and so by induction it follows that M_1 has finite stably free dimension. This concludes the proof of the second case.

Observe that the third statement has been proved in Chapter XX, Lemma 3.8 when A is Noetherian, taking for $\mathfrak{A}$ the abelian category of finite modules, and for $\mathscr{C}$ the family of stably free modules. Mitchell Stokes pointed out to me that the statement is valid in general without Noetherian assumption, and can be proved as follows. We assume that M, M'' have finite free resolutions. We first show that M' is finitely generated. Indeed, suppose first that M is finite free. We have two exact sequences

$$0 \to M' \to M \to M'' \to 0$$
$$0 \to K'' \to F'' \to M'' \to 0$$

where F'' is finite free, and K'' is finitely generated because of the assumption that M'' has a finite free resolution. That M' is finitely generated follows from Schanuel's lemma. If M is not free, one can reduce the finite generation of M' to the case when M is free by a pull-back, which we leave to the reader.

Now suppose that the stably free dimension of M'' is positive. We use the same exact commutative diagram as in the previous cases, with E', E, E'' finite free. The stably free dimension of M''_1 is one less than that of M'', and we are done by induction. This concludes the proof of Theorem 2.7.

This also concludes our general discussion of finite free resolutions. For more information cf. Northcott's book on the subject.

We now come to the second part of this section, which provides an application to polynomial rings.

Theorem 2.8. *Let R be a commutative Noetherian ring. Let x be a variable. If every finite R-module has a finite free resolution, then every finite $R[x]$-module has a finite free resolution.*

In other words, in the category of finite R-modules, if every object is of finite stably free dimension, then the same property applies to the category of finite $R[x]$-modules. Before proving the theorem, we state the application we have in mind.

Theorem 2.9. (Serre). *If k is a field and $x_1, \ldots, x_r$ independent variables, then every finite projective module over $k[x_1, \ldots, x_r]$ is stably free, or equivalently admits a finite free resolution.*

Proof. By induction and Theorem 2.8 we conclude that every finite module over $k[x_1, \ldots, x_r]$ is of finite stably free dimension. (We are using Theorem 2.1.) This concludes the proof.

The rest of this section is devoted to the proof of Theorem 2.8.

Let M be a finite $R[x]$-module. By Chapter X, Corollary 2.8, M has a finite filtration

$$M = M_0 \supset M_1 \supset \cdots \supset M_n = 0$$

such that each factor M_i/M_{i+1} is isomorphic to $R[x]/P_i$ for some prime P_i. In light of Theorem 2.7, it suffices to prove the theorem in case $M = R[x]/P$ where P is prime, which we now assume. In light of the exact sequence

$$0 \to P \to R[x] \to R[x]/P \to 0.$$

and Theorem 2.7, we note that M has a finite free resolution if and only if P does.

Let $\mathfrak{p} = P \cap R$. Then $\mathfrak{p}$ is prime in R. Suppose there is some $M = R[x]/P$ which does not admit a finite free resolution. Among all such M we select one for which the intersection $\mathfrak{p}$ is maximal in the family of prime ideals obtained as above. This is possible in light of one of the basic properties characterizing Noetherian rings.

Let $R_0 = R/\mathfrak{p}$ so R_0 is entire. Let $P_0 = P/\mathfrak{p}R[x]$. Then we may view M as an $R_0[x]$-module, equal to R_0/P_0. Let $f_1, \ldots, f_n$ be a finite set of generators for P_0, and let f be a polynomial of minimal degree in P_0. Let K_0 be the quotient field of R_0. By the euclidean algorithm, we can write

$$f_i = q_i f + r_i \quad \text{for} \quad i = 1, \ldots, n$$

with $q_i, r_i \in K_0[x]$ and $\deg r_i < \deg f$. Let d_0 be a common denominator for the coefficients of all q_i, r_i. Then $d_0 \neq 0$ and

$$d_0 f_i = q_i' f + r_i'$$

where $q_i' = d_0 q_i$ and $r_i' = d_0 r_i$ lie in $R_0[x]$. Since $\deg f$ is minimal in P_0 it follows that $r_i' = 0$ for all i, so

$$d_0 P_0 \subset R_0[x]f = (f).$$

Let $N_0 = P_0/(f)$, so N_0 is a module over $R_0[x]$, and we can also view N_0 as a module over $R[x]$. When so viewed, we denote N_0 by N. Let $d \in R$ be any element reducing to d_0 mod $\mathfrak{p}$. Then $d \notin \mathfrak{p}$ since $d_0 \neq 0$. The module N_0 has a finite filtration such that each factor module of the filtration is isomorphic to some $R_0[x]/Q_0$ where Q_0 is an associated prime of N_0. Let Q be the inverse image of Q_0 in $R[x]$. These prime ideals Q are precisely the associated primes of N in $R[x]$. Since d_0 kills N_0 it follows that d kills N and therefore d lies in every associated prime of N. By the maximality property in the selection of P,

it follows that every one of the factor modules in the filtration of N has a finite free resolution, and by Theorem 2.7 it follows that N itself has a finite free resolution.

Now we view $R_0[x]$ as an $R[x]$-module, via the canonical homomorphism

$$R[x] \to R_0[x] = R[x]/\mathfrak{p}R[x].$$

By assumption, $\mathfrak{p}$ has a finite free resolution as R-module, say

$$0 \to E_n \to \cdots \to E_0 \to \mathfrak{p} \to 0.$$

Then we may simply form the modules $E_i[x]$ in the obvious sense to obtain a finite free resolution of $\mathfrak{p}[x] = \mathfrak{p}R[x]$. From the exact sequence

$$0 \to \mathfrak{p}R[x] \to R[x] \to R_0[x] \to 0$$

we conclude that $R_0[x]$ has a finite free resolution as $R[x]$-module.

Since R_0 is entire, it follows that the principal ideal (f) in $R_0[x]$ is $R[x]$-isomorphic to $R_0[x]$, and therefore has a finite free resolution as $R[x]$-module. Theorem 2.7 applied to the exact sequence of $R[x]$-modules

$$0 \to (f) \to P_0 \to N \to 0$$

shows that P_0 has a finite free resolution; and further applied to the exact sequence

$$0 \to \mathfrak{p}R[x] \to P \to P_0 \to 0$$

shows that P has a finite free resolution, thereby concluding the proof of Theorem 2.8.

§3. UNIMODULAR POLYNOMIAL VECTORS

Let A be a commutative ring. Let $(f_1, \ldots, f_n)$ be elements of A generating the unit ideal. We call such elements **unimodular**. We shall say that they have the **unimodular extension property** if there exists a matrix in $GL_n(A)$ with first column ${}^t(f_1, \ldots, f_n)$. If A is a principal entire ring, then it is a trivial exercise to prove that this is always the case. Serre originally asked the question whether it is true for a polynomial ring $k[x_1, \ldots, x_r]$ over a field k. The problem was solved by Quillen and Suslin. We give here a simplification of Suslin's proof by Vaserstein, also using a previous result of Horrocks. The method is by induction on the number of variables, in some fashion.

We shall write $f = {}^t(f_1, \ldots, f_n)$ for the column vector. We first remark that f has the unimodular extension property if and only if the vector obtained by a permutation of its components has this property. Similarly, we can make

the usual row operations, adding a multiple gf_i to f_j ($j \neq i$), and f has the unimodular extension property if and only if any one of its transforms by row operations has the unimodular extension property.

We first prove the theorem in a context which allows the induction.

Theorem 3.1. (Horrocks). *Let* $(\mathfrak{o}, \mathfrak{m})$ *be a local ring and let* $A = \mathfrak{o}[x]$ *be the polynomial ring in one variable over* $\mathfrak{o}$. *Let* f *be a unimodular vector in* $A^{(n)}$ *such that some component has leading coefficient* 1. *Then* f *has the unimodular extension property.*

Proof. (Suslin). If $n = 1$ or 2 then the theorem is obvious even without assuming that $\mathfrak{o}$ is local. So we assume $n \geqq 3$ and do an induction of the smallest degree d of a component of f with leading coefficient 1. First we note that by the Euclidean algorithm and row operations, we may assume that f_1 has leading coefficient 1, degree d, and that $\deg f_i < d$ for $j \neq 1$. Since f is unimodular, a relation $\sum g_i f_i = 1$ shows that not all coefficients of $f_2, \ldots, f_n$ can lie in the maximal ideal $\mathfrak{m}$. Without loss of generality, we may assume that some coefficient of f_2 does not lie in $\mathfrak{m}$ and so is a unit since $\mathfrak{o}$ is local. Write

$$f_1(x) = x^d + a_{d-1}x^{d-1} + \cdots + a_0 \quad \text{with} \quad a_i \in \mathfrak{o},$$
$$f_2(x) = b_s x^s + \cdots + b_0 \quad \text{with} \quad b_i \in \mathfrak{o},\ s \leqq d - 1,$$

so that some b_i is a unit. Let $\mathfrak{a}$ be the ideal generated by all leading coefficients of polynomials $g_1 f_1 + g_2 f_2$ of degree $\leqq d - 1$. Then $\mathfrak{a}$ contains all the coefficients b_i, $i = 0, \ldots, s$. One sees this by descending induction, starting with b_s which is obvious, and then using a linear combination

$$x^{d-s} f_2(x) - b_s f_1(x).$$

Therefore $\mathfrak{a}$ is the unit ideal, and there exists a polynomial $g_1 f_1 + g_2 f_2$ of degree $\leqq d - 1$ and leading coefficient 1. By row operations, we may now get a polynomial of degree $\leqq d - 1$ and leading coefficient 1 as some component in the i-th place for some $i \neq 1, 2$. Thus ultimately, by induction, we may assume that $d = 0$ in which case the theorem is obvious. This concludes the proof.

Over any commutative ring A, for two column vectors f, g we write $f \sim g$ over A to mean that there exists $M \in GL_n(A)$ such that

$$f = Mg,$$

and we say that f is **equivalent to** g **over** A. Horrocks' theorem states that a unimodular vector f with one component having leading coefficient 1 is $\mathfrak{o}[x]$-equivalent to the first unit vector e^1. We are interested in getting a similar descent over non-local rings. We can write $f = f(x)$, and there is a natural "constant" vector $f(0)$ formed with the constant coefficients. As a corollary of Horrocks' theorem, we get:

Corollary 3.2. *Let $\mathfrak{o}$ be a local ring. Let f be a unimodular vector in $\mathfrak{o}[x]^{(n)}$ such that some component has leading coefficient 1. Then $f \sim f(0)$ over $\mathfrak{o}[x]$.*

Proof. Note that $f(0) \in \mathfrak{o}^{(n)}$ has one component which is a unit. It suffices to prove that over any commutative ring R any element $c \in R^{(n)}$ such that some component is a unit is equivalent over R to e^1, and this is obvious.

Lemma 3.3. *Let R be an entire ring, and let S be a multiplicative subset. Let x, y be independent variables. If $f(x) \sim f(0)$ over $S^{-1}R[x]$, then there exists $c \in S$ such that $f(x + cy) \sim f(x)$ over $R[x, y]$.*

Proof. Let $M \in GL_n(S^{-1}R[x])$ be such that $f(x) = M(x)f(0)$. Then $M(x)^{-1}f(x) = f(0)$ is constant, and thus invariant under translation $x \mapsto x + y$. Let

$$G(x, y) = M(x)M(x + y)^{-1}.$$

Then $G(x, y)f(x + y) = f(x)$. We have $G(x, 0) = I$ whence

$$G(x, y) = I + yH(x, y)$$

with $H(x, y) \in S^{-1}R[x, y]$. There exists $c \in S$ such that cH has coefficients in R. Then $G(x, cy)$ has coefficients in R. Since $\det M(x)$ is constant in $S^{-1}R$, it follows that $\det M(x + cy)$ is equal to this same constant and therefore that $\det G(x, cy) = 1$. This proves the lemma.

Theorem 3.4. *Let R be an entire ring, and let f be a unimodular vector in $R[x]^{(n)}$, such that one component has leading coefficient 1. Then $f(x) \sim f(0)$ over $R[x]$.*

Proof. Let J be the set of elements $c \in R$ such that $f(x + cy)$ is equivalent to $f(x)$ over $R[x, y]$. Then J is an ideal, for if $c \in J$ and $a \in R$ then replacing y by ay in the definition of equivalence shows that $f(x + cay)$ is equivalent to $f(x)$ over $R[x, ay]$, so over $R[x, y]$. Equally easily, one sees that if $c, c' \in J$ then $c + c' \in J$. Now let $\mathfrak{p}$ be a prime ideal of R. By Corollary 3.2 we know that $f(x)$ is equivalent to $f(0)$ over $R_{\mathfrak{p}}[x]$, and by Lemma 3.3 it follows that there exists $c \in R$ and $c \notin \mathfrak{p}$ such that $f(x + cy)$ is equivalent to $f(x)$ over $R[x, y]$. Hence J is not contained in $\mathfrak{p}$, and so J is unit ideal in R, so there exists an invertible matrix $M(x, y)$ over $R[x, y]$ such that

$$f(x + y) = M(x, y)f(x).$$

Since the homomorphic image of an invertible matrix is invertible, we substitute 0 for x in this last relation to conclude the proof of the theorem.

Theorem 3.5. (Quillen-Suslin). *Let k be a field and let f be a unimodular vector in $k[x_1, \ldots, x_r]^{(n)}$. Then f has the unimodular extension property.*

Proof. By induction on r. If $r = 1$ then $k[x_1]$ is a principal ring and the theorem is left to the reader. Assume the theorem for $r - 1$ variables with $r \geqq 2$, and put

$$R = k[x_1, \ldots, x_{r-1}].$$

We view f as a vector of polynomials in the last variable x_r and want to apply Theorem 3.4. We can do so if some component of f has leading coefficient 1 in the variable x_r. We reduce the theorem to this case as follows. The proof of the Noether Normalization Theorem (Chapter VIII, Theorem 2.1) shows that if we let

$$y_r = x_r$$

$$y_i = x_i - x_r^{m_i}$$

then the polynomial vector

$$f(x_1, \ldots, x_r) = g(y_1, \ldots, y_r)$$

has one component with y_r-leading coefficient equal to 1. Hence there exists a matrix $N(y) = M(x)$ invertible over $R[x_r] = R[y_r]$ such that

$$g(y_1, \ldots, y_r) = N(y_1, \ldots, y_r)g(y_1, \ldots, y_{r-1}, 0),$$

and $g(y_1, \ldots, y_{r-1}, 0)$ is unimodular in $k[y_1, \ldots, y_{r-1}]^{(n)}$. We can therefore conclude the proof by induction.

We now give other formulations of the theorem. First we recall that a module E over a commutative ring A is called **stably free** if there exists a finite free module F such that $E \oplus F$ is finite free.

We shall say that a commutative ring A has the **unimodular column extension property** if every unimodular vector $f \in A^{(n)}$ has the unimodular extension property, for all positive integers n.

Theorem 3.6. *Let A be a commutative ring which has the unimodular column extension property. Then every stably free module over A is free.*

Proof. Let E be stably free. We use induction on the rank of the free modules F such that $E \oplus F$ is free. By induction, it suffices to prove that if $E \oplus A$ is free then E is free. Let $E \oplus A = A^{(n)}$ and let

$$p : A^{(n)} \to A$$

be the projection. Let u^1 be a basis of A over itself. Viewing A as a direct summand in $E \oplus A = A^{(n)}$ we write

$$u^1 = {}^t(a_{11}, \ldots, a_{n1}) \quad \text{with} \quad a_{i1} \in A.$$

Then u^1 is unimodular, and by assumption u^1 is the first column of a matrix $M = (a_{ij})$ whose determinant is a unit in A. Let

$$u^j = Me^j \quad \text{for} \quad j = 1, \ldots, n,$$

where e^j is the j-th unit column vector of $A^{(n)}$. Note that u^1 is the first column of M. By elementary column operations, we may change M so that $u^j \in E$ for $j = 2, \ldots, n$. Indeed, if $pe^j = cu^1$ for $j \geqq 2$ we need only replace e^j by $e^j - ce^1$. Without loss of generality we may therefore assume that $u^2, \ldots, u^n$ lie in E. Since M is invertible over A, it follows that M induces an automorphism of $A^{(n)}$ as A-module with itself by

$$X \mapsto MX.$$

It follows immediately from the construction and the fact that $A^{(n)} = E \oplus A$ that M maps the free module with basis $\{e^2, \ldots, e^n\}$ onto E. This concludes the proof.

If we now feed Serre's Theorem 2.9 into the present machinery consisting of the Quillen-Suslin theorem and Theorem 3.6, we obtain the alternative version of the Quillen-Suslin theorem:

Theorem 3.7. *Let k be a field. Then every finite projective module over the polynomial ring $k[x_1, \ldots, x_r]$ is free.*

§4. THE KOSZUL COMPLEX

In this section, we describe a finite complex built out of the alternating product of a free module. This gives an application of the alternating product, and also gives a fundamental construction used in algebraic geometry, both abstract and complex, as the reader can verify by looking at Griffiths-Harris [GrH 78], Chapter V, §3; Grothendieck's [SGA 6]; Hartshorne [Ha 77], Chapter III, §7; and Fulton-Lang [FuL 85], Chapter IV, §2.

We know from Chapter XX that a free resolution of a module allows us to compute certain homology or cohomology groups of a functor. We apply this now to Hom and also to the tensor product. Thus we also get examples of explicit computations of homology, illustrating Chapter XX, by means of the Koszul complex. We shall also obtain a classical application by deriving the so-called Hilbert Syzygy theorem.

Let A be a ring (always assumed commutative) and M a module. A sequence of elements $x_1, \ldots, x_r$ in A is called M-**regular** if $M/(x_1, \ldots, x_r)M \neq 0$, if x_1

is not divisor of zero in M, and for $i \geqq 2$, x_i is not divisor of 0 in

$$M/(x_1, \ldots, x_{i-1})M.$$

It is called **regular** when $M = A$.

Proposition 4.1. *Let $I = (x_1, \ldots, x_r)$ be generated by a regular sequence in A. Then I/I^2 is free of dimension r over A/I.*

Proof. Let $\bar{x}_i$ be the class of x_i mod I^2. It suffices to prove that $\bar{x}_1, \ldots, \bar{x}_r$ are linearly independent. We do this by induction on r. For $r = 1$, if $\bar{a}\bar{x} = 0$, then $ax = bx^2$ for some $b \in A$, so $x(a - bx) = 0$. Since x is not zero divisor in A, we have $a = bx$ so $\bar{a} = 0$.

Now suppose the proposition true for the regular sequence $x_1, \ldots, x_{r-1}$. Suppose

$$\sum_{i=1}^{r} \bar{a}_i \bar{x}_i = 0 \quad \text{in} \quad I/I^2.$$

We may assume that $\sum a_i x_i = 0$ in A; otherwise $\sum a_i x_i = \sum y_i x_i$ with $y_i \in I$ and we can replace a_i by $a_i - y_i$ without changing $\bar{a}_i$.

Since x_r is not zero divisor in $A/(x_1, \ldots, x_{r-1})$ there exist $b_i \in A$ such that

$$a_r x_r + \sum_{i=1}^{r-1} a_i x_i = 0 \Rightarrow a_r = \sum_{i=1}^{r-1} b_i x_i \Rightarrow \sum_{i=1}^{r-1} (a_i + b_i x_r) x_i = 0.$$

By induction,

$$a_j + b_j x_r \in \sum_{i=1}^{r-1} A x_i \qquad (j = 1, \ldots, r-1)$$

so $a_j \in I$ for all j, so $\bar{a}_j = 0$ for all j, thus proving the proposition.

Let K, L be complexes, which we write as direct sums

$$K = \bigoplus K_p \quad \text{and} \quad L = \bigoplus L_q$$

with $p, q \in \mathbf{Z}$. Usually, $K_p = L_q = 0$ for $p, q < 0$. Then the **tensor product** $K \otimes L$ is the complex such that

$$(K \otimes L)_n = \bigoplus_{p+q=n} K_p \otimes L_q;$$

and for $u \in K_p$, $v \in L_q$ the differential is defined by

$$d(u \otimes v) = du \otimes v + (-1)^p u \otimes dv.$$

(Carry out the detailed verification, which is routine, that this gives a complex.)

Let A be a commutative ring and $x \in A$. We define the complex $K(x)$ to have $K_0(x) = A$, $K_1(x) = Ae_1$, where e_1 is a symbol, Ae_1 is the free module of rank 1 with basis $\{e_1\}$, and the boundary map is defined by $de_1 = x$, so the complex can be represented by the sequence

$$\begin{array}{ccccccc} 0 & \longrightarrow & Ae_1 & \xrightarrow{d} & A & \longrightarrow & 0 \\ & & \| & & \| & & \\ 0 & \longrightarrow & K_1(x) & \longrightarrow & K_0(x) & \longrightarrow & 0 \end{array}$$

More generally, for elements $x_1, \ldots, x_r \in A$ we define the **Koszul complex** $K(x) = K(x_1, \ldots, x_r)$ as follows. We put:

$K_0(x) = A$;

$K_1(x) =$ free module E with basis $\{e_1, \ldots, e_r\}$;

$K_p(x) =$ free module $\bigwedge^p E$ with basis $\{e_{i_1} \wedge \cdots \wedge e_{i_p}\}$, $i_1 < \cdots < i_p$;

$K_r(x) =$ free module $\bigwedge^r E$ of rank 1 with basis $e_1 \wedge \cdots \wedge e_r$.

We define the **boundary maps** by $de_i = x_i$ and in general

$$d : K_p(x) \to K_{p-1}(x)$$

by

$$d(e_{i_1} \wedge \cdots \wedge e_{i_p}) = \sum_{j=1}^{p} (-1)^{j-1} x_{i_j} e_{i_1} \wedge \cdots \wedge \widehat{e_{i_j}} \wedge \cdots \wedge e_{i_p}.$$

A direct verification shows that $d^2 = 0$, so we have a complex

$$0 \to K_r(x) \to \cdots \to K_p(x) \to \cdots \to K_1(x) \to A \to 0$$

The next lemma shows the extent to which the complex is independent of the ideal $I = (x_1, \ldots, x_r)$ generated by (x). Let

$$I = (x_1, \ldots, x_r) \supset I' = (y_1, \ldots, y_r)$$

be two ideals of A. We have a natural ring homomorphism

$$\text{can} : A/I' \to A/I.$$

Let $\{e'_1, \ldots, e'_r\}$ be a basis for $K_1(y)$, and let

$$y_i = \sum c_{ij} x_j \quad \text{with} \quad c_{ij} \in A.$$

We **define** $f_1 : K_1(y) \to K_1(x)$ by

$$f_1 e'_i = \sum c_{ij} e_j$$

and

$$f_p = f_1 \wedge \cdots \wedge f_1, \qquad \text{product taken } p \text{ times.}$$

Let $D = \det(c_{ij})$ be the determinant. Then for $p = r$ we get that

$$f_r : K_r(y) \to K_r(x) \text{ is multiplication by } D.$$

Lemma 4.2. *Notation as above, the homomorphisms f_p define a morphism of Koszul complexes:*

$$\begin{array}{ccccccccccccccc}
0 & \longrightarrow & K_r(y) & \longrightarrow \cdots \longrightarrow & K_p(y) & \longrightarrow \cdots \longrightarrow & K_1(y) & \longrightarrow & A & \longrightarrow & A/I' & \longrightarrow & 0 \\
 & & \downarrow {\scriptstyle f_r = D} & & \downarrow {\scriptstyle f_p} & & \downarrow {\scriptstyle f_1} & & \downarrow {\scriptstyle \mathrm{id}} & & \downarrow {\scriptstyle \mathrm{can}} & & \\
0 & \longrightarrow & K_r(x) & \longrightarrow \cdots \longrightarrow & K_p(x) & \longrightarrow \cdots \longrightarrow & K_1(x) & \longrightarrow & A & \longrightarrow & A/I & \longrightarrow & 0
\end{array}$$

and define an isomorphism if D is a unit in A, for instance if (y) is a permutation of (x).

Proof. By definition

$$f(e'_{i_1} \wedge \cdots \wedge e'_{i_p}) = \left(\sum_{j=1}^{r} c_{i_1 j} e_j\right) \wedge \cdots \wedge \left(\sum_{j=1}^{r} c_{i_p j} e_j\right).$$

Then

$$\begin{aligned}
& fd(e'_{i_1} \wedge \cdots \wedge e'_{i_p}) \\
&= f\left(\sum_k (-1)^{k-1} y_{i_k} e'_{i_1} \wedge \cdots \wedge \widehat{e'_{i_k}} \wedge \cdots \wedge e'_{i_p}\right) \\
&= \sum_k (-1)^{k-1} y_{i_k} \left(\sum_{j=1}^{r} c_{i_1 j} e_j\right) \wedge \cdots \wedge \widehat{\sum_k} \wedge \cdots \wedge \left(\sum_{j=1}^{r} c_{i_p j} e_j\right) \\
&= \sum (-1)^{k-1} \left(\sum_{j=1}^{r} c_{i_1 j} e_j\right) \wedge \cdots \wedge \underbrace{\left(\sum_{j=1}^{r} c_{i_k j} x_j e_j\right)}_{\text{omitted}} \wedge \cdots \wedge \left(\sum_{j=1}^{r} c_{i_p j} e_j\right) \\
&= df(e'_{i_1} \wedge \cdots \wedge e'_{i_p})
\end{aligned}$$

using $y_{i_k} = \sum c_{i_k j} x_j$. This concludes the proof that the f_p define a homomorphism of complexes.

In particular, if (x) and (y) generate the same ideal, and the determinant D is a unit (i.e. the linear transformation going from (x) to (y) is invertible over the ring), then the two Koszul complexes are isomorphic.

The next lemma gives us a useful way of making inductions later.

Proposition 4.3. *There is a natural isomorphism*

$$K(x_1, \dots, x_r) \approx K(x_1) \otimes \cdots \otimes K(x_r).$$

Proof. The proof will be left as an exercise.

Let $I = (x_1, \dots, x_r)$ be the ideal generated by $x_1, \dots, x_r$. Then directly from the definitions we see that the 0-th homology of the Koszul complex is simply A/IA.

More generally, let M be an A-module. Define the **Koszul complex of** M by

$$K(x; M) = K(x_1, \dots, x_r; M) = K(x_1, \dots, x_r) \otimes_A M$$

Then this complex looks like

$$0 \to K_r(x) \otimes M \to \cdots \to K_2(x) \otimes_A M \to M^{(r)} \to M \to 0.$$

We sometimes abbreviate $H_p(x; M)$ for $H_p K(x; M)$. The first and last homology groups are then obtained directly from the definition of boundary. We get

$$H_0(K(x; M)) \approx M/IM;$$

$$H_r(K(x); M) = \{v \in M \text{ such that } x_i v = 0 \text{ for all } i = 1, \dots, r\}.$$

In light of Proposition 4.3, we study generally what happens to a tensor product of any complex with $K(x)$, when x consists of a single element. Let $y \in A$ and let C be an arbitrary complex of A-modules. We have an exact sequence of complexes

(1) $$0 \to C \to C \otimes K(y) \to (C \otimes K(y))/C \to 0$$

made explicit as follows.

$$\begin{array}{ccccccccc}
 & & \downarrow & & \downarrow & & \downarrow & & \\
0 & \to & C_{n+1} & \to & (C_{n+1} \otimes A) \oplus (C_n \otimes K_1(y)) & \to & C_n \otimes K_1(y) & \to & 0 \\
 & & \downarrow & & \downarrow & & \downarrow {\scriptstyle d_n \otimes \mathrm{id}} & & \\
0 & \to & C_n & \to & (C_n \otimes A) \oplus (C_{n-1} \otimes K_1(y) & \to & C_{n-1} \otimes K_1(y)) & \to & 0 \\
 & & \downarrow & & \downarrow & & \downarrow {\scriptstyle d_{n-1} \otimes \mathrm{id}} & & \\
0 & \to & C_{n-1} & \to & (C_{n-1} \otimes A) \oplus (C_{n-2} \otimes K_1(y)) & \to & C_{n-2} \otimes K_1(y) & \to & 0 \\
 & & \downarrow & & \downarrow & & \downarrow & &
\end{array}$$

We note that $C \otimes K_1(y)$ is just C with a dimension shift by one unit, in other words

(2) $$(C \otimes K_1(y))_{n+1} = C_n \otimes K_1(y).$$

In particular,

(3) $$H_{n+1}(C \otimes K(y)/C) \approx H_n(C).$$

Associated with an exact sequence of complexes, we have the homology sequence, which in this case yields the long exact sequence

$$\longrightarrow H_{n+1}(C) \longrightarrow H_{n+1}(C \otimes K_1(y))$$

$$\longrightarrow H_{n+1}(C \otimes K(y)/C) \xrightarrow{\partial} H_n(C)$$

$$H_{n+1}(C \otimes K(y)/C) \approx H_n(C)$$

which we write stacked up according to the index:

(4) $$\begin{aligned} &\to H_{p+1}(C) \to H_{p+1}(C) \to H_{p+1}(C \otimes K(y)) \to \\ &\to H_p(C) \to H_p(C) \to H_p(C \otimes K(y)) \to \end{aligned}$$

ending in lowest dimension with

(5) $$\to H_1(C) \to H_1(C \otimes K(y)) \to H_0(C) \to H_0(C).$$

Furthermore, a direct application of the definition of the boundary map and the tensor product of complexes yields:

The boundary map on $H_p(C)$ ($p \geqq 0$) is induced by multiplication by $(-1)^p y$:

(6) $$\partial = (-1)^p m(y) : H_p(C) \to H_p(C).$$

Indeed, write

$$(C \otimes K(y))_p = (C_p \otimes A) \oplus (C_{p-1} \otimes K_1(y)) \approx C_p \oplus C_{p-1}.$$

Let $(v, w) \in C_p \oplus C_{p-1}$ with $v \in C_p$ and $w \in C_{p-1}$. Then directly from the definitions,

(7) $$d(v, w) = (dv + (-1)^{p-1}yw, dw).$$

To see (6), one merely follows up the definitions of the boundary, taking an element $w \in C_p \approx C_p \otimes K_1(y)$, lifting back to $(0, w)$, applying d, and lifting back to C_p. If we start with a cycle, i.e. $dw = 0$, then the map is well defined on the homology class, with values in the homology.

Lemma 4.4. *Let $y \in A$ and let C be a complex as above. Then $m(y)$ annihilates $H_p(C \otimes K(y))$ for all $p \geqq 0$.*

Proof. If (v, w) is a cycle, i.e. $d(v, w) = 0$, then from (7) we get at once that $(yv, yw) = d(0, (-1)^p v)$, which proves the lemma.

In the applications we have in mind, we let $y = x_r$ and

$$C = K(x_1, \ldots, x_{r-1}; M) = K(x_1, \ldots, x_{r-1}) \otimes M.$$

Then we obtain:

Theorem 4.5.(a) *There is an exact sequence with maps as above:*

$$\to H_pK(x_1, \ldots, x_{r-1}; M) \to H_pK(x_1, \ldots, x_{r-1}; M) \to H_pK(x_1, \ldots, x_r; M)$$
$$\cdots \to H_1(x_1, \ldots, x_r; M) \to H_0(x_1, \ldots, x_{r-1}; M) \xrightarrow{m(x_r)} H_0(x_1, \ldots, x_{r-1}; M).$$

(b) *Every element of $I = (x_1, \ldots, x_r)$ annihilates $H_p(x; M)$ for $p \geqq 0$.*
(c) *If $I = A$, then $H_p(x; M) = 0$ for all $p \geqq 0$.*

Proof. This is immediate from Proposition 4.3 and Lemma 4.4.

We define the **augmented Koszul complex** to be

$$0 \to K_r(x; M) \to \cdots \to K_1(x; M) = M^{(r)} \to M \to M/IM \to 0.$$

Theorem 4.6. *Let M be an A-module.*
(a) *Let $x_1, \ldots, x_r$ be a regular sequence for M. Then $H_pK(x; M) = 0$ for $p > 0$. (Of course, $H_0K(x; M) = M/IM$.) In other words, the augmented Koszul complex is exact.*
(b) *Conversely, suppose A is local, and $x_1, \ldots, x_r$ lie in the maximal ideal of A. Suppose M is finite over A, and also assume that $H_1K(x; M) = 0$. Then $(x_1, \ldots, x_r)$ is M-regular.*

Proof. We prove (a) by induction on r. If $r = 1$ then $H_1(x; M) = 0$ directly from the definition. Suppose $r > 1$. We use the exact sequence of Theorem 4.5(a). If $p > 1$ then $H_p(x; M)$ is between two homology groups which are 0, so $H_p(x; M) = 0$. If $p = 1$, we use the very end of the exact sequence of Theorem 4.5(a), noting that $m(x_r)$ is injective, so by induction we find $H_1(x; M) = 0$ also, thus proving (a).

As to (b), by Lemma 4.4 and the hypothesis, we get an exact sequence

$$H_1(x_1, \ldots, x_{r-1}; M) \xrightarrow{m(x_r)} H_1(x_1, \ldots, x_{r-1}; M) \to H_1(x; M) = 0,$$

so $m(x_r)$ is surjective. By Nakayama's lemma, it follows that

$$H_1(x_1, \ldots, x_{r-1}; M) = 0.$$

By induction $(x_1, \ldots, x_{r-1})$ is an M-regular sequence. Looking again at the tail end of the exact sequence as in (a) shows that x_r is $M/(x_1, \ldots, x_{r-1})M$-regular, whence proving (b) and the theorem.

We note that (b), which uses only the triviality of H_1 (and not all H_p) is due to Northcott [No 68], 8.5, Theorem 8. By (a), it follows that $H_p = 0$ for $p > 0$.

An important special case of Theorem 4.6(a) is when $M = A$, in which case we restate the theorem in the form:

Let $x_1, \ldots, x_r$ be a regular sequence in A. Then $K(x_1, \ldots, x_r)$ is a free resolution of A/I:

$$0 \to K_r(x) \to \cdots \to K_1(x) \to A \to A/I \to 0.$$

In particular, A/I has Tor-*dimension $\leqq r$.*

For the Hom functor, we have:

Theorem 4.7. *Let $x_1, \ldots, x_r$ be a regular sequence in A. Then there is an isomorphism*

$$\varphi_{x,M}: H^r(\mathrm{Hom}(K(x), M)) \to M/IM$$

to be described below.

Proof. The module $K_r(x)$ is 1-dimensional, with basis $e_1 \wedge \cdots \wedge e_r$. Depending on this basis, we have an isomorphism

$$\mathrm{Hom}(K_r(x), M) \approx M,$$

whereby a homomorphism is determined by its value at the basis element in M. Then directly from the definition of the boundary map d_r in the Koszul complex, which is

$$d_r : e_1 \wedge \cdots \wedge e_r \mapsto \sum_{j=1}^{r} (-1)^{j-1} x_j e_1 \wedge \cdots \wedge \hat{e}_j \wedge \cdots \wedge e_r$$

we see that

$$\begin{aligned} H^r(\mathrm{Hom}(K_r(x), M) &\approx \mathrm{Hom}(K_r(x), M)/d^{r-1}\,\mathrm{Hom}(K_{r-1}(x), M) \\ &\approx M/IM. \end{aligned}$$

This proves the theorem.

The reader who has read Chapter XX knows that the i-th homology group of $\mathrm{Hom}(K(x), M)$ is called $\mathrm{Ext}^i(A/I, M)$, determined up to a unique isomorphism by the complex, since two resolutions of A/I differ by a morphism of complexes, and two such morphisms differ by a homotopy which induces a homology isomorphism. Thus Theorem 4.7 gives an isomorphism

$$\varphi_{x,M} : \mathrm{Ext}^r(A/I, M) \to M/IM.$$

In fact, we shall obtain morphisms of the Koszul complex from changing the sequence. We go back to the hypothesis of Lemma 4.2.

Lemma 4.8. *If $I = (x) = (y)$ where (x), (y) are two regular sequences, then we have a commutative diagram*

$$\begin{array}{ccc} & & M/IM \\ & \nearrow^{\varphi_{x,M}} & \\ \operatorname{Ext}^r(A/I, M) & & \downarrow D = \det(c_{ij}) \\ & \searrow_{\varphi_{y,M}} & \\ & & M/IM \end{array}$$

where all the maps are isomorphisms of A/I-modules.

The fact that we are dealing with A/I-modules is immediate since multiplication by an element of A commutes with all homomorphisms in sight, and I annihilates A/I.

By Proposition 4.1, we know that I/I^2 is a free module of rank r over A/I. Hence

$$\bigwedge^r(I/I^2)$$

is a free module of rank 1, with basis $\bar{x}_1 \wedge \cdots \wedge \bar{x}_r$ (where the bar denotes residue class mod I^2). Taking the dual of this exterior product, we see that under a change of basis, it transforms according to the inverse of the determinant mod I^2. This allows us to get a canonical isomorphism as in the next theorem.

Theorem 4.9. *Let $x_1, \ldots, x_r$ be a regular sequence in A, and let $I = (x)$. Let M be an A-module. Let*

$$\psi_{x,M} : M/IM \to (M/IM) \otimes \bigwedge^r(I/I^2)^{\text{dual}}$$

be the embedding determined by the basis $(\bar{x}_1 \wedge \cdots \wedge \bar{x}_r)^{\text{dual}}$ of $\bigwedge^r(I/I^2)^{\text{dual}}$. Then the composite isomorphism

$$\operatorname{Ext}^r(A/I, M) \xrightarrow{\varphi_{x,M}} M/IM \xrightarrow{\psi_{x,M}} (M/IM) \otimes \bigwedge^r(I/I^2)^{\text{dual}}$$

is a functorial isomorphism, independent of the choice of regular generators for I.

We also have the analogue of Theorem 4.5 in intermediate dimensions.

Theorem 4.10. *Let $x_1, \ldots, x_r$ be an M-regular sequence in A. Let $I = (x)$. Then*

$$\operatorname{Ext}^i(A/I, M) = 0 \quad \textit{for} \quad i < r.$$

Proof. For the proof, we assume that the reader is acquainted with the exact homology sequence. Assume by induction that $\operatorname{Ext}^i(A/I, M) = 0$ for

$i < r - 1$. Then we have the exact sequence

$$0 = \mathrm{Ext}^{i-1}(A/I, M/x_1M) \to \mathrm{Ext}^i(A/I, M) \xrightarrow{x_1} \mathrm{Ext}^i(A/I, M)$$

for $i < r$. But $x_1 \in I$ so multiplication by x_1 induces 0 on the homology groups, which gives $\mathrm{Ext}^i(A/I, M) = 0$ as desired.

Let $L_N \to N \to 0$ be a free resolution of a module N. By definition,

$$\mathrm{Tor}_i^A(N, M) = i\text{-th homology of the complex } L \otimes M.$$

This is independent of the choice of L_N up to a unique isomorphism. We now want to do for Tor what we have just done for Ext.

Theorem 4.11. *Let $I = (x_1, \ldots, x_r)$ be an ideal of A generated by a regular sequence of length r.*

(i) *There is a natural isomorphism*

$$\mathrm{Tor}_i^A(A/I, A/I) \approx \textstyle\bigwedge_{A/I}^i(I/I^2), \quad \text{for} \quad i \geqq 0.$$

(ii) *Let L be a free A/I-module, extended naturally to an A-module. Then*

$$\mathrm{Tor}_i^A(L, A/I) \approx L \otimes \textstyle\bigwedge_{A/I}^i(I/I^2), \quad \text{for} \quad i \geqq 0.$$

These isomorphisms will follow from the next considerations.

First we use again that the residue classes $\bar{x}_1, \ldots, \bar{x}_r$ mod I^2 form a basis of I/I^2 over A/I. Therefore we have a unique isomorphism of complexes

$$\varphi_x : K(x) \otimes A/I \to \textstyle\bigwedge(I/I^2) = \bigoplus \bigwedge^i(I/I^2)$$

with zero differentials on the right-hand side, such that

$$e_{i_1} \wedge \cdots \wedge e_{i_p} \mapsto \bar{x}_{i_1} \wedge \cdots \wedge \bar{x}_{i_p}.$$

Lemma 4.12. *Let $I = (x) \supset I' = (y)$ be two ideals generated by regular sequences of length r. Let $f : K(y) \to K(x)$ be the morphism of Koszul complexes defined in Lemma 4.2. Then the following diagram is commutative:*

$$\begin{array}{ccc} K(y) \otimes A/I' & \xrightarrow{\varphi_y} & \bigwedge_{A/I'}(I'/I'^2) \\ \Big\downarrow{\scriptstyle f \otimes \mathrm{can}} & & \Big\downarrow{\scriptstyle \text{canonical hom}} \\ K(x) \otimes A/I & \xrightarrow[\varphi_x]{} & \bigwedge_{A/I}(I/I^2) \end{array}$$

Proof. We have

$$\begin{aligned}\varphi_x \circ (f \otimes \text{can})(e'_{i_1} \wedge \cdots \wedge e'_{i_p} \otimes 1) \\ = \sum_{j=2}^{r} c_{i_1 j}\bar{x}_j \wedge \cdots \wedge \sum_{j=1}^{r} c_{i_p j}\bar{x}_j \\ = \bar{y}_{i_1} \wedge \cdots \wedge \bar{y}_{i_p} = \text{can}(\varphi_y(e'_{i_1} \wedge \cdots \wedge e'_{i_p})).\end{aligned}$$

This proves the lemma.

In particular, if $I' = I$ then we have the commutative diagram

$$\begin{array}{ccc} K(y) & & \\ {\scriptstyle f\otimes d}\downarrow & \searrow^{\varphi_y} & \\ & & \bigwedge(I/I^2) \\ & \nearrow_{\varphi_x} & \\ K(x) & & \end{array}$$

which shows that the identification of $\text{Tor}_i(A/I, A/I)$ with $\bigwedge^i(I/I^2)$ via the choices of bases is compatible under one isomorphism of the Koszul complexes, which provide a resolution of A/I. Since any other homomorphism of Koszul complexes is homotopic to this one, it follows that this identification does not depend on the choices made and proves the first part of Theorem 4.11.

The second part follows at once, because we have

$$\begin{aligned}\text{Tor}_i^A(A/I, L) = H_i(K(x) \otimes L) = H_i((K(x) \otimes_A A/I) \otimes_{A/I} L \\ = \textstyle\bigwedge^i_{A/I}(I/I^2) \otimes L.\end{aligned}$$

This concludes the proof of Theorem 4.11.

Example. Let k be a field and let $A = k[x_1, \ldots, x_r]$ be the polynomial ring in r variables. Let $I = (x_1, \ldots, x_r)$ be the ideal generated by the variables. Then $A/I = k$, and therefore Theorem 4.11 yields for $i \geqq 0$:

$$\text{Tor}_i^A(k, k) \approx \textstyle\bigwedge^i_k(I/I^2)$$

$$\text{Tor}_i^A(L, k) \approx L \otimes \textstyle\bigwedge^i_k(I/I^2)$$

Note that in the present case, we can think of I/I^2 as the vector space over k with basis $\bar{x}_1, \ldots, \bar{x}_r$. Then A can be viewed as the symmetric algebra SE, where E is this vector space. We can give a specific example of the Koszul complex in this context as in the next theorem, given for a free module.

Theorem 4.13. *Let E be a finite free module of rank r over the ring R. For each $p = 1, \ldots, r$ there is a unique homomorphism*

$$d_p : \bigwedge^p E \otimes SE \to \bigwedge^{p-1} E \otimes SE$$

such that

$$d_i((x_1 \wedge \cdots \wedge x_p) \otimes y)$$

$$= \sum_{i=1}^{p} (-1)^{i-1} (x_1 \wedge \cdots \wedge \widehat{x_i} \wedge \cdots \wedge x_p) \otimes (x_i \otimes y)$$

where $x_i \in E$ and $y \in SE$. This gives the resolution

$$0 \to \bigwedge^r E \otimes SE \to \bigwedge^{r-1} E \otimes SE \to \cdots \to \bigwedge^0 E \otimes SE \to R \to 0$$

Proof. The above definitions are merely examples of the Koszul complex for the symmetric algebra SE with respect to the regular sequence consisting of some basis of E.

Since d_p maps $\bigwedge^p E \otimes S^q E$ into $\bigwedge^{p-1} E \otimes S^{q+1} E$, we can decompose this complex into a direct sum corresponding to a given graded component, and hence:

Corollary 4.14. *For each integer $n \geqq 1$, we have an exact sequence*

$$0 \to \bigwedge^r E \otimes S^{n-r} E \to \cdots \to \bigwedge^1 E \otimes S^{n-1} E \to S^n E \to 0$$

where $S^j E = 0$ for $j < 0$.

Finally, we give an application to a classical theorem of Hilbert. The polynomial ring $A = k[x_1, \ldots, x_r]$ is naturally graded, by the degrees of the homogeneous components. *We shall consider graded modules, where the grading is in dimensions $\geqq 0$, and we assume that homomorphisms are graded of degree* 0.

So suppose M is a graded module (and thus $M_i = 0$ for $i < 0$) and M is finite over A. Then we can find a graded surjective homomorphism

$$L_0 \to M \to 0$$

where L_0 is finite free. Indeed, let $w_1, \ldots, w_n$ be homogeneous generators of M. Let $e_1, \ldots, e_n$ be basis elements for a free module L_0 over A. We give L_0 the grading such that if $a \in A$ is homogeneous of degree d then ae_i is homogeneous of degree

$$\deg ae_i = \deg a + \deg w_i.$$

Then the homomorphism of L_0 onto M sending $e_i \mapsto w_i$ is graded as desired.

The kernel M_1 is a graded submodule of L_0. Repeating the process, we can find a surjective homomorphism

$$L_1 \to M_1 \to 0.$$

We continue in this way to obtain a graded resolution of M. We want this resolution to stop, and the possibility of its stopping is given by the next theorem.

Theorem 4.15. (Hilbert Syzygy Theorem). *Let k be a field and*

$$A = k[x_1, \ldots, x_r]$$

the polynomial ring in r variables. Let M be a graded module over A, and let

$$0 \to K \to L_{r-1} \to \cdots \to L_0 \to M \to 0$$

be an exact sequence of graded homomorphisms of graded modules, such that $L_0, \ldots, L_{r-1}$ are free. Then K is free. If M is in addition finite over A and $L_0, \ldots, L_{r-1}$ are finite free, then K is finite free.

Proof. From the Koszul complex we know that $\mathrm{Tor}_i(M, k) = 0$ for $i > r$ and all M. By dimension shifting, it follows that

$$\mathrm{Tor}_i(K, k) = 0 \quad \text{for} \quad i > 0.$$

The theorem is then a consequence of the next result.

Theorem 4.16. *Let F be a graded finite module over $A = k[x_1, \ldots, x_r]$. If $\mathrm{Tor}_1(F, k) = 0$ then F is free.*

Proof. The method is essentially to do a Nakayama type argument in the case of the non-local ring A. First note that

$$F \otimes k = F/IF$$

where $I = (x_1, \ldots, x_r)$. Thus $F \otimes k$ is naturally an $A/I = k$-module. Let $v_1, \ldots, v_n$ be homogeneous elements of F whose residue classes mod IF form a basis of F/IF over k. Let L be a free module with basis $e_1, \ldots, e_n$. Let

$$L \to F$$

be the graded homomorphism sending $e_i \mapsto v_i$ for $i = 1, \ldots, n$. It suffices to prove that this is an isomorphism. Let C be the cokernel, so we have the exact sequence

$$L \to F \to C \to 0.$$

Tensoring with k yields the exact sequence

$$L \otimes k \to F \otimes k \to C \otimes k \to 0.$$

Since by construction the map $L \otimes k \to F \otimes k$ is surjective, it follows that $C \otimes k = 0$. But C is graded, so the next lemma shows that $C = 0$.

Lemma 4.17. *Let N be a graded module over $A = k[x_1, \ldots, x_r]$. Let $I = (x_1, \ldots, x_r)$. If $N/IN = 0$ then $N = 0$.*

Proof. This is immediate by using the grading, looking at elements of N of smallest degree if they exist, and using the fact that elements of I have degree > 0.

We now get an exact sequence of graded modules

$$0 \to E \to L \to F \to 0$$

and we must show that $E = 0$. But the exact homology sequence and our assumption yields

$$0 = \mathrm{Tor}_1(F, k) \to E \otimes k \to L \otimes k \to F \otimes k \to 0.$$

By construction $L \otimes k \to F \otimes k$ is an isomorphism, and hence $E \otimes k = 0$. Lemma 4.17 now shows that $E = 0$. This concludes the proof of the syzygy theorem.

Remark. The only place in the proof where we used that k is a field is in the proof of Theorem 4.16 when we picked homogeneous elements $v_1, \ldots, v_n$ in M whose residue classes mod IM form a basis of M/IM over A/IA. Hilbert's theorem can be generalized by making the appropriate hypothesis which allows us to carry out this step, as follows.

Theorem 4.18. *Let R be a commutative local ring and let $A = R[x_1, \ldots, x_r]$ be the polynomial ring in r variables. Let M be a graded finite module over A, projective over R. Let*

$$0 \to K \to L_{r-1} \to \cdots \to L_0 \to M \to 0$$

be an exact sequence of graded homomorphisms of graded modules such that $L_0, \ldots, L_{r-1}$ are finite free. Then K is finite free.

Proof. Replace k by R everywhere in the proof of the Hilbert syzygy theorem. We use the fact that a finite projective module over a local ring is free. Not a word needs to be changed in the above proof with the following exception. We note that the projectivity propagates to the kernels and cokernels in the given resolution. Thus F in the statement of Theorem 4.16 may be assumed projective, and each graded component is projective. Then F/IF is projective over $A/IA = R$, and so is each graded component. Since a finite projective module over a local ring is free, and one gets the freeness by lifting a basis from the residue class field, we may pick $v_1, \ldots, v_n$ homogeneous exactly as we did in the proof of Theorem 4.16. This concludes the proof.

EXERCISES

For exercises 1 through 4 on the Koszul complex, see [No 68], Chapter 8.

1. Let $0 \to M' \to M \to M'' \to 0$ be an exact sequence of A-modules. Show that tensoring with the Koszul complex $K(x)$ one gets an exact sequence of complexes, and therefore an exact homology sequence

$$0 \to H_r K(x; M') \to H_r K(x; M) \to H_r K(x; M'') \to \cdots$$
$$\cdots \to H_p K(x; M') \to H_p K(x; M) \to H_p K(x; M'') \to \cdots$$
$$\cdots \to H_0 K(x; M') \to H_0 K(x; M) \to H_0 K(x; M'') \to 0$$

2. (a) Show that there is a unique homomorphism of complexes

$$f: K(x; M) \to K(x_1, \ldots, x_{r-1}; M)$$

such that for $v \in M$:

$$f_p(e_{i_1} \wedge \cdots \wedge e_{i_p} \otimes v) = \begin{cases} e_{i_1} \wedge \cdots \wedge e_{i_p} \otimes x_r v & \text{if } i_p = r \\ e_{i_1} \wedge \cdots \wedge e_{i_p} \otimes v & \text{if } i_p = r. \end{cases}$$

(b) Show that f is injective if x_r is not a divisor of zero in M.
(c) For a complex C, denote by $C(-1)$ the complex shifted by one place to the left, so $C(-1)_n = C_{n-1}$ for all n. Let $\overline{M} = M/x_r M$. Show that there is a unique homomorphism of complexes

$$g: K(x_1, \ldots, x_{r-1}, 1; M) \to K(x_1, \ldots, x_{r-1}; \overline{M})(-1)$$

such that for $v \in M$:

$$g_p(e_{i_1} \wedge \cdots \wedge c_{i_p} \otimes v) = \begin{cases} e_{i_1} \wedge \cdots \wedge e_{i_{p-1}} \otimes v & \text{if } i_p = r \\ 0 & \text{if } i_p < r. \end{cases}$$

(d) If x_r is not a divisor of 0 in M, show that the following sequence is exact:

$$0 \to K(x; M) \xrightarrow{f} K(x_1, \ldots, x_{r-1}, 1; M) \xrightarrow{g} K(x_1, \ldots, x_{r-1}; \overline{M})(-1) \to 0.$$

Using Theorem 4.5(c), conclude that for all $p \geqq 0$, there is an isomorphism

$$H_p K(x; M) \xrightarrow{\approx} H_p K(x_1, \ldots, x_{r-1}; \overline{M}).$$

3. Assume A and M Noetherian. Let I be an ideal of A. Let $a_1, \ldots, a_k$ be an M-regular sequence in I. Show that this sequence can be extended to a maximal M-regular sequence $a_1, \ldots, a_q$ in I, in other words an M-regular sequence such that there is no M-regular sequence $a_1, \ldots, a_{q+1}$ in I.

4. Again assume A and M Noetherian. Let $I = (x_1, \ldots, x_r)$ and let $a_1, \ldots, a_q$ be a maximal M-regular sequence in I. Assume $IM \neq M$. Prove that

$$H_{r-q}(x; M) \neq 0 \text{ but } H_p(x; M) = 0 \text{ for } p > r - q.$$

[See [No 68], 8.5 Theorem 6. The result is similar to the result in Exercise 5, and generalizes Theorem 4.5(a). See also [Mat 80], pp. 100-103. The result shows that

all maximal M-regular sequences in M have the same length, which is called the I-**depth** of M and is denoted by $\operatorname{depth}_I(M)$. For the proof, let s be the maximal integer such that $H_sK(x; M) \neq 0$. By assumption, $H_0(x; M) = M/IM \neq 0$, so s exists. We have to prove that $q + s = r$. First note that if $q = 0$ then $s = r$. Indeed, if $q = 0$ then every element of I is zero divisor in M, whence I is contained in the union of the associated primes of M, whence in some associated prime of M. Hence $H_r(x; M) \neq 0$.

Next assume $q > 0$ and proceed by induction. Consider the exact sequence

$$0 \to M \xrightarrow{a_1} M \to M/a_1M \to 0$$

where the first map is $m(a_1)$. Since I annihilates $H_p(x; M)$ by Theorem 4.5(c), we get an exact sequence

$$0 \to H_p(x; M) \to H_p(x; M/a_1M) \to H_{p-1}(x; M) \to 0.$$

Hence $H_{s+1}(x; M/a_1M) \neq 0$, but $H_p(x; M/a_1M) = 0$ for $p \geqq s + 2$. From the hypothesis that $a_1, \ldots, a_q$ is a maximal M-regular sequence, it follows at once that $a_2, \ldots, a_q$ is maximal M/a_1M-regular in I, so by induction, $q - 1 = r - (s + 1)$ and hence $q + s = r$, as was to be shown.]

5. The following exercise combines some notions of Chapter XX on homology, and some notions covered in this chapter and in Chapter X, §5. Let M be an A-module.

Let A be Noetherian, M finite module over A, and I an ideal of A such that $IM \neq M$. Let r be an integer $\geqq 1$. Prove that the following conditions are equivalent:

(i) $\operatorname{Ext}^i(N, M) = 0$ for all $i < r$ and all finite modules N such that $\operatorname{supp}(N) \subset \mathfrak{X}(I)$.

(ii) $\operatorname{Ext}^i(A/I, M) = 0$ for all $i < r$.

(iii) There exists a finite module N with $\operatorname{supp}(N) = \mathfrak{X}(I)$ such that

$$\operatorname{Ext}^i(N, M) = 0 \quad \text{for all } i < r.$$

(iv) There exists an M-regular sequence $a_1, \ldots, a_r$ in I.

[*Hint*: (i) $\Rightarrow$ (ii) $\Rightarrow$ (iii) is clear. For (iii) $\Rightarrow$ (iv), first note that

$$0 = \operatorname{Ext}^0(N, M) = \operatorname{Hom}(N, M).$$

Assume $\operatorname{supp}(N) = \mathfrak{X}(I)$. Find an M-regular element in I. If there is no such element, then I is contained in the set of divisors of 0 of M in A, which is the union of the associated primes. Hence $I \subset P$ for some associated prime P. This yields an injection $A/P \subset M$, so

$$0 \neq \operatorname{Hom}_{A_P}(A_P/PA_P, M).$$

By hypothesis, $N_P \neq 0$ so $N_P/PN_P \neq 0$, and N_P/PN_P is a vector space over A_P/PA_P, so there exists a non-zero A_P/PA_P homomorphism

$$N_P/PN_P \to M_P,$$

so $\operatorname{Hom}_{A_P}(N_P, M_P) \neq 0$, whence $\operatorname{Hom}(N, M) \neq 0$, a contradiction. This proves the existence of one regular element a_1.

Now let $M_1 = M/a_1M$. The exact sequence

$$0 \to M \xrightarrow{a_1} M \to M/a_1M \to 0$$

yields the exact cohomology sequence

$$\to \operatorname{Ext}^i(N, M) \to \operatorname{Ext}^i(N, M/a_1M) \to \operatorname{Ext}^{i+1}(N, M) \to$$

so $\operatorname{Ext}^i(N, M/a_1M) = 0$ for $i < r - 1$. By induction there exists an M_1-regular sequence $a_2, \dots, a_r$ and we are done.

Last, (iv) $\Rightarrow$ (i). Assume the existence of the regular sequence. By induction, $\operatorname{Ext}^i(N, a_1M) = 0$ for $i < r - 1$. We have an exact sequence for $i < r$:

$$0 \to \operatorname{Ext}^i(N, M) \xrightarrow{a_1} \operatorname{Ext}^i(N, M)$$

But $\operatorname{supp}(N) = \mathfrak{X}(\operatorname{ann}(N)) \subset \mathfrak{X}(I)$, so $I \subset \operatorname{rad}(\operatorname{ann}(N))$, so a_1 is nilpotent on N. Hence a_1 is nilpotent on $\operatorname{Ext}^i(N, M)$, so $\operatorname{Ext}^i(N, M) = 0$. Done.] See Matsumura's [Mat 70], p. 100, Theorem 28. The result is useful in algebraic geometry, with for instance $M = A$ itself. One thinks of A as the affine coordinate ring of some variety, and one thinks of the equations $a_i = 0$ as defining hypersurface sections of this variety, and the simultaneous equations $a_1 = \cdots = a_r = 0$ as defining a complete intersection. The theorem gives a cohomological criterion in terms of Ext for the existence of such a complete intersection.

APPENDIX 1

The Transcendence of e and π

The proof which we shall give here follows the classical method of Gelfond and Schneider, properly formulated. It is based on a theorem concerning values of functions satisfying differential equations, and it had been recognized for some time that such values are subject to severe restrictions, in various contexts. Here, we deal with the most general algebraic differential equation.

We shall assume that the reader is acquainted with elementary facts concerning functions of a complex variable. Let f be an entire function (i.e. a function which is holomorphic on the complex plane). For our purposes, we say f is of order $\leqq \rho$ if there exists a number $C > 1$ such that for all large R we have

$$|f(z)| \leqq C^{R^\rho}$$

whenever $|z| \leqq R$. A meromorphic function is said to be of order $\leqq \rho$ if it is a quotient of entire functions of order $\leqq \rho$.

Theorem. *Let K be a finite extension of the rational numbers. Let $f_1, \dots, f_N$ be meromorphic functions of order $\leqq \rho$. Assume that the field $K(f_1, \dots, f_N)$ has transcendence degree $\geqq 2$ over K, and that the derivative $D = d/dz$ maps the ring $K[f_1, \dots, f_N]$ into itself. Let $w_1, \dots, w_m$ be distinct complex numbers not lying among the poles of the f_i, such that*

$$f_i(w_\nu) \in K$$

for all $i = 1, \dots, N$ and $\nu = 1, \dots, m$. Then $m \leqq 10\rho[K : \mathbf{Q}]$.

Corollary 1. (Hermite-Lindemann). *If α is algebraic (over $\mathbf{Q}$) and $\neq 0$, then e^α is transcendental. Hence π is transcendental.*

Proof. Suppose that α and e^α are algebraic. Let $K = \mathbf{Q}(\alpha, e^\alpha)$. The two functions z and e^z are algebraically independent over K (trivial), and the ring $K[z, e^z]$ is obviously mapped into itself by the derivative. Our functions take on algebraic values in K at $\alpha, 2\alpha, \ldots, m\alpha$ for any m, contradiction. Since $e^{2\pi i} = 1$, it follows that $2\pi i$ is transcendental.

Corollary 2. (Gelfond-Schneider). *If α is algebraic $\neq 0, 1$ and if β is algebraic irrational, then $\alpha^\beta = e^{\beta \log \alpha}$ is transcendental.*

Proof. We proceed as in Corollary 1, considering the functions $e^{\beta t}$ and e^t which are algebraically independent because β is assumed irrational. We look at the numbers $\log \alpha, 2 \log \alpha, \ldots, m \log \alpha$ to get a contradiction as in Corollary 1.

Before giving the main arguments proving the theorem, we state some lemmas. The first two, due to Siegel, have to do with integral solutions of linear homogeneous equations.

Lemma 1. *Let*

$$a_{11}x_1 + \cdots + a_{1n}x_n = 0$$
$$\cdots$$
$$a_{r1}x_1 + \cdots + a_{rn}x_n = 0$$

be a system of linear equations with integer coefficients a_{ij}, and $n > r$. Let A be a number such that $|a_{ij}| \leqq A$ for all i, j. Then there exists an integral, non-trivial solution with

$$|x_j| \leqq 2(2nA)^{r/(n-r)}.$$

Proof. We view our system of linear equations as a linear equation $L(X) = 0$, where L is a linear map, $L: \mathbf{Z}^{(n)} \to \mathbf{Z}^{(r)}$, determined by the matrix of coefficients. If B is a positive number, we denote by $\mathbf{Z}^{(n)}(B)$ the set of vectors X in $\mathbf{Z}^{(n)}$ such that $|X| \leqq B$ (where $|X|$ is the maximum of the absolute values of the coefficients of X). Then L maps $\mathbf{Z}^{(n)}(B)$ into $\mathbf{Z}^{(r)}(nBA)$. The number of elements in $\mathbf{Z}^{(n)}(B)$ is $\geqq B^n$ and $\leqq (2B+1)^n$. We seek a value of B such that there will be two distinct elements X, Y in $\mathbf{Z}^{(n)}(B)$ having the same image, $L(X) = L(Y)$. For this, it will suffice that $B^n > (2nBA)^r$, and thus it will suffice that

$$B = (2nA)^{r/(n-r)}.$$

We take $X - Y$ as the solution of our problem.

Let K be a finite extension of $\mathbf{Q}$, and let I_K be the integral closure of $\mathbf{Z}$ in K. From Exercise 5 of Chapter IX, we know that I_K is a free module over $\mathbf{Z}$, of dimension $[K:\mathbf{Q}]$. We view K as contained in the complex numbers. If

$\alpha \in K$, a conjugate of α will be taken to be an element $\sigma\alpha$, where σ is an embedding of K in $\mathbf{C}$. By the **size** of a set of elements of K we shall mean the maximum of the absolute values of all conjugates of these elements.

By the size of a vector $X = (x_1, \ldots, x_n)$ we shall mean the size of the set of its coordinates.

Let $\omega_1, \ldots, \omega_M$ be a basis of I_K over $\mathbf{Z}$. Let $\alpha \in I_K$, and write

$$\alpha = a_1\omega_1 + \cdots + a_M\omega_M.$$

Let $\omega'_1, \ldots, \omega'_M$ be the dual basis of $\omega_1, \ldots, \omega_M$ with respect to the trace. Then we can express the (Fourier) coefficients a_j of α as a trace,

$$a_j = \mathrm{Tr}(\alpha\omega'_j).$$

The trace is a sum over the conjugates. Hence the size of these coefficients is bounded by the size of α, times a fixed constant, depending on the size of the elements ω'_j.

Lemma 2. *Let K be a finite extension of $\mathbf{Q}$. Let*

$$\alpha_{11}x_1 + \cdots + \alpha_{1n}x_n = 0$$
$$\cdots$$
$$\alpha_{r1}x_1 + \cdots + \alpha_{rn}x_n = 0$$

be a system of linear equations with coefficients in I_K, and $n > r$. Let A be a number such that $\mathrm{size}(\alpha_{ij}) \leqq A$, *for all i, j. Then there exists a non-trivial solution X in I_K such that*

$$\mathrm{size}(X) \leqq C_1(C_2nA)^{r/(n-r)},$$

where C_1, C_2 are constants depending only on K.

Proof. Let $\omega_1, \ldots, \omega_M$ be a basis of I_K over $\mathbf{Z}$. Each x_j can be written

$$x_j = \xi_{j1}\omega_1 + \cdots + \xi_{jM}\omega_M$$

with unknowns $\xi_{j\lambda}$. Each α_{ij} can be written

$$\alpha_{ij} = a_{ij1}\omega_1 + \cdots + a_{ijM}\omega_M$$

with integers $a_{ij\lambda} \in \mathbf{Z}$. If we multiply out the $\alpha_{ij}x_j$, we find that our linear equations with coefficients in I_K are equivalent to a system of rM linear equations in the nM unknowns $\xi_{j\lambda}$, with coefficients in $\mathbf{Z}$, whose size is bounded by CA, where C is a number depending only on M and the size of the elements ω_λ, together with the products $\omega_\lambda\omega_\mu$, in other words where C depends only on K. Applying Lemma 1, we obtain a solution in terms of the $\xi_{j\lambda}$, and hence a solution X in I_K, whose size satisfies the desired bound.

The next lemma has to do with estimates of derivatives. By the size of a polynomial with coefficients in K, we shall mean the size of its set of coefficients. A **denominator** for a set of elements of K will be any positive rational integer whose product with every element of the set is an algebraic integer. We define in a similar way a denominator for a polynomial with coefficients in K. We abbreviate "denominator" by den.

Let

$$P(T_1, \dots, T_N) = \sum \alpha_{(\nu)} M_{(\nu)}(T)$$

be a polynomial with complex coefficients, and let

$$Q(T_1, \dots, T_N) = \sum \beta_{(\nu)} M_{(\nu)}(T)$$

be a polynomial with real coefficients $\geqq 0$. We say that Q **dominates** P if $|\alpha_{(\nu)}| \leq \beta_{(\nu)}$ for all (ν). It is then immediately verified that the relation of dominance is preserved under addition, multiplication, and taking partial derivatives with respect to the variables $T_1, \dots, T_N$.

Lemma 3. *Let K be of finite degree over $\mathbf{Q}$. Let $f_1, \dots, f_N$ be functions, holomorphic on a neighborhood of a point $w \in \mathbf{C}$, and assume that $D = d/dz$ maps the ring $K[f_1, \dots, f_N]$ into itself. Assume that $f_i(w) \in K$ for all i. Then there exists a number C_1 having the following property. Let $P(T_1, \dots, T_N)$ be a polynomial with coefficients in K, of degree $\leqq r$. If we set $f = P(f_1, \dots, f_N)$, then we have, for all positive integers k,*

$$\text{size}(D^k f(w)) \leqq \text{size}(P) r^k k! C_1^{k+r}$$

Furthermore, there is a denominator for $D^k f(w)$ bounded by $\text{den}(P)C_1^{k+r}$.

Proof. There exist polynomials $P_i(T_1, \dots, T_N)$ with coefficients in K such that

$$Df_i = P_i(f_1, \dots, f_N).$$

Let h be the maximum of their degrees. There exists a unique derivation $\bar{D}$ on $K[T_1, \dots, T_N]$ such that $\bar{D}T_i = P_i(T_1, \dots, T_N)$. For any polynomial P we have

$$\bar{D}(P(T_1, \dots, T_N)) = \sum_{i=1}^{N} (D_i P)(T_1, \dots, T_N) \cdot P_i(T_1, \dots, T_N),$$

where $D_1, \dots, D_N$ are the partial derivatives. The polynomial P is dominated by

$$\text{size}(P)(1 + T_1 + \cdots + T_N)^r,$$

and each P_i is dominated by $\text{size}(P_i)(1 + T_1 + \cdots + T_N)^h$. Thus $\bar{D}P$ is dominated by

$$\text{size}(P) C_2 r (1 + T_1 + \cdots + T_N)^{r+h}.$$

Proceeding inductively, one sees that $\bar{D}^k P$ is dominated by

$$\text{size}(P) C_3^k r^k k! (1 + T_1 \cdots + T_N)^{r+kh}.$$

Substituting values $f_i(w)$ for T_i, we obtain the desired bound on $D^k f(w)$. The second assertion concerning denominators is proved also by a trivial induction.

We now come to the main part of the proof of our theorem. Let f, g be two functions among $f_1, \ldots, f_N$ which are algebraically independent over K. Let r be a positive integer divisible by $2m$. We shall let r tend to infinity at the end of the proof.

Let

$$F = \sum_{i,j=1}^{r} b_{ij} f^i g^j$$

have coefficients b_{ij} in K. Let $n = r^2/2m$. We can select the b_{ij} not all equal to 0, and such that

$$D^k F(w_v) = 0$$

for $0 \leqq k < n$ and $v = 1, \ldots, m$. Indeed, we have to solve a system of mn linear equations in $r^2 = 2mn$ unknowns. Note that

$$\frac{mn}{2mn - mn} = 1.$$

We multiply these equations by a denominator for the coefficients. Using the estimate of Lemma 3, and Lemma 2, we can in fact take the b_{ij} to be algebraic integers, whose size is bounded by

$$O(r^n n! C_1^{n+r}) \leqq O(n^{2n})$$

for $n \to \infty$.

Since f, g are algebraically independent over K, our function F is not identically zero. We let s be the smallest integer such that all derivatives of F up to order $s - 1$ vanish at all points $w_1, \ldots, w_m$, but such that $D^s F$ does not vanish at one of the w, say w_1. Then $s \geqq n$. We let

$$\gamma = D^s F(w_1) \neq 0.$$

Then γ is an element of K, and by Lemma 3, it has a denominator which is bounded by $O(C_1^s)$ for $s \to \infty$. Let c be this denominator. The norm of $c\gamma$ from K to $\mathbf{Q}$ is then a non-zero rational integer. Each conjugate of $c\gamma$ is bounded by $O(s^{5s})$. Consequently, we get

$$1 \leqq |N_{\mathbf{Q}}^K(c\gamma)| \leqq O(s^{5s})^{[K:\mathbf{Q}]-1} |\gamma|, \tag{1}$$

where $|\gamma|$ is the fixed absolute value of γ, which will now be estimated very well by global arguments.

Let θ be an entire function of order $\leqq \rho$, such that θf and θg are entire, and $\theta(w_1) \neq 0$. Then $\theta^{2r}F$ is entire. We consider the entire function

$$H(z) = \frac{\theta(z)^{2r}F(z)}{\prod_{\nu=1}^{m}(z - w_\nu)^s}.$$

Then $H(w_1)$ differs from $D^sF(w_1)$ by obvious factors, bounded by $C_4^s s!$. By the maximum modulus principle, its absolute value is bounded by the maximum of H on a large circle of radius R. If we take R large, then $z - w_\nu$ has approximately the same absolute value as R, and consequently, on the circle of radius R, $H(z)$ is bounded in absolute value by an expression of type

$$\frac{s^{3s}C_5^{2rR^\rho}}{R^{ms}}.$$

We select $R = s^{1/2\rho}$. We then get the estimate

$$|\gamma| \leqq \frac{s^{4s}C_6^s}{s^{ms/2\rho}}.$$

We now let r tend to infinity. Then both n and s tend to infinity. Combining this last inequality with inequality (1), we obtain the desired bound on m. This concludes the proof.

Of course, we made no effort to be especially careful in the powers of s occurring in the estimates, and the number 10 can obviously be decreased by exercising a little more care in the estimates.

The theorem we proved is only the simplest in an extensive theory dealing with problems of transcendence degree. In some sense, the theorem is best possible without additional hypotheses. For instance, if $P(t)$ is a polynomial with integer coefficients, then $e^{P(t)}$ will take the value 1 at all roots of P, these being algebraic. Furthermore, the functions

$$t, e^t, e^{t^2}, \ldots, e^{t^n}$$

are algebraically independent, but take on values in $\mathbf{Q}(e)$ for all integral values of t.

However, one expects rather strong results of algebraic independence to hold. Lindemann proved that *if* $\alpha_1, \ldots, \alpha_n$ *are algebraic numbers, linearly independent over* $\mathbf{Q}$, *then*

$$e^{\alpha_1}, \ldots, e^{\alpha_n}$$

are algebraically independent.

More generally, Schanuel has made the following conjecture: If $\alpha_1, \ldots, \alpha_n$ are complex numbers, linearly independent over $\mathbf{Q}$, then the transcendence degree of

$$\alpha_1, \ldots, \alpha_n, e^{\alpha_1}, \ldots, e^{\alpha_n}$$

should be $\geqq n$.

From this one would deduce at once the algebraic independence of e and π (looking at $1, 2\pi i, e, e^{2\pi i}$), and all other independence statements concerning the ordinary exponential function and logarithm which one feels to be true, for instance, the statement that π cannot lie in the field obtained by starting with the algebraic numbers, adjoining values of the exponential function, taking algebraic closure, and iterating these two operations. Such statements have to do with values of the exponential function lying in certain fields of transcendence degree $< n$, and one hopes that by a suitable deepening of Theorem 1, one will reach the desired results.

APPENDIX 2

Some Set Theory

§1. DENUMERABLE SETS

Let n be a positive integer. Let J_n be the set consisting of all integers k, $1 \leqq k \leqq n$. If S is a set, we say that S has n elements if there is a bijection between S and J_n. Such a bijection associates with each integer k as above an element of S, say $k \mapsto a_k$. Thus we may use J_n to "count" S. Part of what we assume about the basic facts concerning positive integers is that if S has n elements, then the integer n is uniquely determined by S.

One also agrees to say that a set has 0 elements if the set is empty.

We shall say that a set S is **denumerable** if there exists a bijection of S with the set of positive integers $\mathbf{Z}^+$. Such a bijection is then said to **enumerate** the set S. It is a mapping

$$n \mapsto a_n$$

which to each positive integer n associates an element of S, the mapping being injective and surjective.

If D is a denumerable set, and $f : S \to D$ is a bijection of some set S with D, then S is also denumerable. Indeed, there is a bijection $g : D \to \mathbf{Z}^+$, and hence $g \circ f$ is a bijection of S with $\mathbf{Z}^+$.

Let T be a set. A **sequence** of elements of T is simply a mapping of $\mathbf{Z}^+$ into T. If the map is given by the association $n \mapsto x_n$, we also write the sequence as $\{x_n\}_{n \geqq 1}$, or also $\{x_1, x_2, \ldots\}$. For simplicity, we also write $\{x_n\}$ for the sequence. Thus we think of the sequence as prescribing a first, second, ..., n-th element of T. We use the same braces for sequences as for sets, but the context will always make our meaning clear.

Examples. The even positive integers may be viewed as a sequence $\{x_n\}$ if we put $x_n = 2n$ for $n = 1, 2, \ldots$. The odd positive integers may also be viewed as a sequence $\{y_n\}$ if we put $y_n = 2n - 1$ for $n = 1, 2, \ldots$. In each case, the sequence gives an enumeration of the given set.

We also use the word *sequence* for mappings of the natural numbers into a set, thus allowing our sequences to start from 0 instead of 1. If we need to specify whether a sequence starts with the 0-th term or the first term, we write

$$\{x_n\}_{n \geqq 0} \quad \text{or} \quad \{x_n\}_{n \geqq 1}$$

according to the desired case. Unless otherwise specified, however, we always assume that a sequence will start with the first term. Note that from a sequence $\{x_n\}_{n \geqq 0}$ we can define a new sequence by letting $y_n = x_{n-1}$ for $n \geqq 1$. Then $y_1 = x_0, y_2 = x_1, \ldots$. Thus there is no essential difference between the two kinds of sequences.

Given a sequence $\{x_n\}$, we call x_n the n-th term of the sequence. A sequence may very well be such that all its terms are equal. For instance, if we let $x_n = 1$ for all $n \geqq 1$, we obtain the sequence $\{1, 1, 1, \ldots\}$. Thus there is a difference between a sequence of elements in a set T, and a subset of T. In the example just given, the set of all terms of the sequence consists of one element, namely the single number 1.

Let $\{x_1, x_2, \ldots\}$ be a sequence in a set S. By a **subsequence** we shall mean a sequence $\{x_{n_1}, x_{n_2}, \ldots\}$ such that $n_1 < n_2 < \cdots$. For instance, if $\{x_n\}$ is the sequence of positive integers, $x_n = n$, the sequence of even positive integers $\{x_{2n}\}$ is a subsequence.

An enumeration of a set S is of course a sequence in S.

A set is **finite** if the set is empty, or if the set has n elements for some positive integer n. If a set is not finite, it is called **infinite**.

Occasionally, a map of J_n into a set T will be called a **finite sequence** in T. A finite sequence is written as usual,

$$\{x_1, \ldots, x_n\} \quad \text{or} \quad \{x_i\}_{i=1,\ldots,n}.$$

When we need to specify the distinction between finite sequences and maps of $\mathbf{Z}^+$ into T, we call the latter infinite sequences. Unless otherwise specified, we shall use the word sequence to mean infinite sequence.

Proposition 1.1. *Let D be an infinite subset of $\mathbf{Z}^+$. Then D is denumerable, and in fact there is a unique enumeration of D, say $\{k_1, k_2, \ldots\}$ such that*

$$k_1 < k_2 < \cdots < k_n < k_{n+1} < \cdots.$$

Proof. We let k_1 be the smallest element of D. Suppose inductively that we have defined $k_1 < \cdots < k_n$, in such a way that any element k in D which is not equal to $k_1, \ldots, k_n$ is $> k_n$. We define k_{n+1} to be the smallest element of D which is $> k_n$. Then the map $n \mapsto k_n$ is the desired enumeration of D.

Corollary 1.2. *Let S be a denumerable set and D an infinite subset of S. Then D is denumerable.*

Proof. Given an enumeration of S, the subset D corresponds to a subset of $\mathbf{Z}^+$ in this enumeration. Using Proposition 1.1, we conclude that we can enumerate D.

Proposition 1.3. *Every infinite set contains a denumerable subset.*

Proof. Let S be an infinite set. For every non-empty subset T of S, we select a definite element a_T in T. We then proceed by induction. We let x_1 be the chosen element a_S. Suppose that we have chosen $x_1, \ldots, x_n$ having the property that for each $k = 2, \ldots, n$ the element x_k is the selected element in the subset which is the complement of $\{x_1, \ldots, x_{k-1}\}$. We let x_{n+1} be the selected element in the complement of the set $\{x_1, \ldots, x_n\}$. By induction, we thus obtain an association $n \mapsto x_n$ for all positive integers n, and since $x_n \neq x_k$ for all $k < n$ it follows that our association is injective, i.e. gives an enumeration of a subset of S.

Proposition 1.4. *Let D be a denumerable set, and $f: D \to S$ a surjective mapping. Then S is denumerable or finite.*

Proof. For each $y \in S$, there exists an element $x_y \in D$ such that $f(x_y) = y$ because f is surjective. The association $y \mapsto x_y$ is an injective mapping of S into D, because if

$$y, z \in S \quad \text{and} \quad x_y = x_z$$

then

$$y = f(x_y) = f(x_z) = z.$$

Let $g(y) = x_y$. The image of g is a subset of D and D is denumerable. Since g is a bijection between S and its image, it follows that S is denumerable or finite.

Proposition 1.5. *Let D be a denumerable set. Then $D \times D$ (the set of all pairs (x, y) with $x, y \in D$) is denumerable.*

Proof. There is a bijection between $D \times D$ and $\mathbf{Z}^+ \times \mathbf{Z}^+$, so it will suffice to prove that $\mathbf{Z}^+ \times \mathbf{Z}^+$ is denumerable. Consider the mapping of $\mathbf{Z}^+ \times \mathbf{Z}^+ \to \mathbf{Z}^+$ given by

$$(m, n) \mapsto 2^n 3^m.$$

It is injective, and by Proposition 1.1, our result follows.

Proposition 1.6. *Let $\{D_1, D_2, \ldots\}$ be a sequence of denumerable sets. Let S be the union of all sets D_i $(i = 1, 2, \ldots)$. Then S is denumerable.*

Proof. For each $i = 1, 2, \dots$ we enumerate the elements of D_i, as indicated in the following notation:

$$\begin{array}{ll} D_1\colon & \{x_{11}, x_{12}, x_{13}, \dots\} \\ D_2\colon & \{x_{21}, x_{22}, x_{23}, \dots\} \\ & \quad \dots \\ D_i\colon & \{x_{i1}, x_{i2}, x_{i3}, \dots\} \\ & \quad \dots \end{array}$$

The map $f\colon \mathbf{Z}^+ \times \mathbf{Z}^+ \to D$ given by

$$f(i, j) = x_{ij}$$

is then a surjective map of $\mathbf{Z}^+ \times \mathbf{Z}^+$ onto S. By Proposition 1.4, it follows that S is denumerable.

Corollary 1.7. *Let F be a non-empty finite set and D a denumerable set. Then $F \times D$ is denumerable. If $S_1, S_2, \dots$ are a sequence of sets, each of which is finite or denumerable, then the union $S_1 \cup S_2 \cup \cdots$ is denumerable or finite.*

Proof. There is an injection of F into $\mathbf{Z}^+$ and a bijection of D with $\mathbf{Z}^+$. Hence there is an injection of $F \times \mathbf{Z}^+$ into $\mathbf{Z}^+ \times \mathbf{Z}^+$ and we can apply Corollary 1.2 and Proposition 1.6 to prove the first statement. One could also define a surjective map of $\mathbf{Z}^+ \times \mathbf{Z}^+$ onto $F \times D$. (Cf. Exercises 1 and 4.) As for the second statement, each finite set is contained in some denumerable set, so that the second statement follows from Proposition 1.1 and 1.6.

For convenience, we shall say that a set is **countable** if it is either finite or denumerable.

§2. ZORN'S LEMMA

In order to deal efficiently with infinitely many sets simultaneously, one needs a special property. To state it, we need some more terminology.

Let S be a set. An **ordering** (also called partial ordering) of S is a relation, written $x \leqq y$, among some pairs of elements of S, having the following properties.

ORD 1. *We have $x \leqq x$.*

ORD 2. *If $x \leqq y$ and $y \leqq z$ then $x \leqq z$.*

ORD 3. *If $x \leqq y$ and $y \leqq x$ then $x = y$.*

We sometimes write $y \geqq x$ for $x \leqq y$. Note that we don't require that the relation $x \leqq y$ or $y \leqq x$ hold for every pair of elements (x, y) of S. Some pairs may not be comparable. If the ordering satisfies this additional property, then we say that it is a **total ordering**.

Example 1. Let G be a group. Let S be the set of subgroups. If H, H' are subgroups of G, we define

$$H \leqq H'$$

if H is a subgroup of H'. One verifies immediately that this relation defines an ordering on S. Given two subgroups H, H' of G, we do not necessarily have $H \leqq H'$ or $H' \leqq H$.

Example 2. Let R be a ring, and let S be the set of left ideals of R. We define an ordering in S in a way similar to the above, namely if L, L' are left ideals of R, we define

$$L \leqq L'$$

if $L \subset L'$.

Example 3. Let X be a set, and S the set of subsets of X. If Y, Z are subsets of X, we define $Y \leqq Z$ if Y is a subset of Z. This defines an ordering on S.

In all these examples, the relation of ordering is said to be that of inclusion.

In an ordered set, if $x \leqq y$ and $x \neq y$ we then write $x < y$.

Let A be an ordered set, and B a subset. Then we can define an ordering on B by defining $x \leqq y$ for $x, y \in B$ to hold if and only if $x \leqq y$ in A. We shall say that R_0 is the ordering on B **induced** by R, or is the **restriction** to B of the partial ordering of A.

Let S be an ordered set. By a **least** element of S (or a **smallest** element) one means an element $a \in S$ such that $a \leqq x$ for all $x \in S$. Similarly, by a **greatest element** one means an element b such that $x \leqq b$ for all $x \in S$.

By a **maximal element** m of S one means an element such that if $x \in S$ and $x \geqq m$, then $x = m$. Note that a maximal element need not be a greatest element. There may be many maximal elements in S, whereas if a greatest element exists, then it is unique (proof?).

Let S be an ordered set. We shall say that S is **totally ordered** if given $x, y \in S$ we have necessarily $x \leqq y$ or $y \leqq x$.

Example 4. The integers **Z** are totally ordered by the usual ordering. So are the real numbers.

Let S be an ordered set, and T a subset. An **upper bound** of T (in S) is an element $b \in S$ such that $x \leqq b$ for all $x \in T$. A **least upper bound** of T in S is an upper bound b such that, if c is another upper bound, then $b \leqq c$. We shall say

that S is **inductively ordered** if every non-empty totally ordered subset has an upper bound.

We shall say that S is **strictly** inductively ordered if every non-empty totally ordered subset has a least upper bound.

In Examples 1, 2, 3, in each case, the set is strictly inductively ordered. To prove this, let us take Example 2. Let T be a non-empty totally ordered subset of the set of subgroups of G. This means that if $H, H' \in T$, then $H \subset H'$ or $H' \subset H$. Let U be the union of all sets in T. Then:

1. U is a subgroup. *Proof*: If $x, y \in U$, there exist subgroups $H, H' \in T$ such that $x \in H$ and $y \in H'$. If, say, $H \subset H'$, then both $x, y \in H'$ and hence $xy \in H'$. Hence $xy \in U$. Also, $x^{-1} \in H'$, so $x^{-1} \in U$. Hence U is a subgroup.

2. U is an upper bound for each element of T. *Proof*: Every $H \in T$ is contained in U, so $H \leqq U$ for all $H \in T$.

3. U is a least upper bound for T. *Proof*: Any subgroup of G which contains all the subgroups $H \in T$ must then contain their union U.

The proof that the sets in Examples 2, 3 are strictly inductively ordered is entirely similar.

We can now state the property mentioned at the beginning of the section.

Zorn's Lemma. *Let S be a non-empty inductively ordered set. Then there exists a maximal element in S.*

As an example of Zorn's lemma, we shall now prove the infinite version of a theorem given in Chapters 1, §7, and XIV, §2, namely:

Let R be an entire, principal ring and let E be a free module over R. Let F be a submodule. Then F is free. In fact, if $\{v_i\}_{i \in I}$ is a basis for E, and $F \neq \{0\}$, then there exists a basis for F indexed by a subset of I.

Proof. For each subset J of I we let E_J be the free submodule of E generated by all $v_j, j \in J$, and we let $F_J = E_J \cap F$. We let S be the set of all pairs (F_J, w) where J is a subset of I, and $w : J' \to F_J$ is a basis of F_J indexed by a subset J' of J. We write w_j instead of $w(j)$ for $j \in J'$. If (F_J, w) and (F_K, u) are such pairs, we define $(F_J, w) \leqq (F_K, u)$ if $J \subset K$, if $J' \subset K'$, and if the restriction of u to J is equal to w. (In other words, the basis u for F_K is an extension of the basis w for F_J.) This defines an ordering on S, and it is immediately verified that S is in fact inductively ordered, and non-empty (say by the finite case of the result). We can therefore apply Zorn's lemma. Let (F_J, w) be a maximal element. We contend that $J = I$ (this will prove our result). Suppose $J \neq I$ and let $k \in I$ but $k \notin J$. Let $K = J \cup \{k\}$. If

$$E_{J \cup \{k\}} \cap F = F_J,$$

then (F_K, w) is a bigger pair than (F_J, w) contradicting the maximality assumption. Otherwise there exist elements of F_K which can be written in the form

$$cv_k + y$$

with some $y \in E_J$ and $c \in R$, $c \neq 0$. The set of all elements $c \in R$ such that there exists $y \in E_J$ for which $cv_k + y \in F$ is an ideal. Let a be a generator of this ideal, and let

$$w_k = av_k + y$$

be an element of F, with $y \in E_J$. If $z \in F_K$ then there exists $b \in R$ such that $z - bw_k \in E_J$. But $z - bw_k \in F$, whence $z - bw_k \in F_J$. It follows at once that the family consisting of w_j $(j \in J)$ and w_k is a basis for F_K, thus contradicting the maximality again. This proves what we wanted.

Zorn's lemma could be just taken as an axiom of set theory. However, it is not psychologically completely satisfactory as an axiom, because its statement is too involved, and one does not visualize easily the existence of the maximal element asserted in that statement. We show how one can prove Zorn's lemma from other properties of sets which everyone would immediately grant as acceptable psychologically.

From now on to the end of the proof of Theorem 2.1, we let A be a non-empty partially ordered and strictly inductively ordered set. We recall that **strictly inductively ordered** means that every nonempty totally ordered subset has a least upper bound. We assume given a map $f: A \to A$ such that for all $x \in A$ we have $x \leqq f(x)$. We could call such a map an **increasing** map.

Let $a \in A$. Let B be a subset of A. We shall say that B is **admissible** if:

1. B contains a.
2. We have $f(B) \subset B$.
3. Whenever T is a non-empty totally ordered subset of B, the least upper bound of T in A lies in B.

Then B is also strictly inductively ordered, by the induced ordering of A. We shall prove:

Theorem 2.1. (Bourbaki). *Let A be a non-empty partially ordered and strictly inductively ordered set. Let $f: A \to A$ be an increasing mapping. Then there exists an element $x_0 \in A$ such that $f(x_0) = x_0$.*

Proof. Suppose that A were totally ordered. By assumption, it would have a least upper bound $b \in A$, and then

$$b \leqq f(b) \leqq b,$$

so that in this case, our theorem is clear. The whole problem is to reduce the theorem to that case. In other words, what we need to find is a totally ordered admissible subset of A.

If we throw out of A all elements $x \in A$ such that x is not $\geqq a$, then what remains is obviously an admissible subset. Thus without loss of generality, we may assume that A has a least element a, that is $a \leqq x$ for all $x \in A$.

Let M be the intersection of all admissible subsets of A. Note that A itself is an admissible subset, and that all admissible subsets of A contain a, so that M is not empty. Furthermore, M is itself an admissible subset of A. To see this, let $x \in M$. Then x is in every admissible subset, so $f(x)$ is also in every admissible subset, and hence $f(x) \in M$. Hence $f(M) \subset M$. If T is a totally ordered non-empty subset of M, and b is the least upper bound of T in A, then b lies in every admissible subset of A, and hence lies in M. It follows that M is the smallest admissible subset of A, and that any admissible subset of A contained in M is equal to M.

We shall prove that M is totally ordered, and thereby prove Theorem 2.1.

[First we make some remarks which don't belong to the proof, but will help in the understanding of the subsequent lemmas. Since $a \in M$, we see that $f(a) \in M$, $f \circ f(a) \in M$, and in general $f^n(a) \in M$. Furthermore,

$$a \leqq f(a) \leqq f^2(a) \leqq \cdots.$$

If we had an equality somewhere, we would be finished, so we may assume that the inequalities hold. Let D_0 be the totally ordered set $\{f^n(a)\}_{n \geqq 0}$. Then D_0 looks like this:

$$a < f(a) < f^2(a) < \cdots < f^n(a) < \cdots.$$

Let a_1 be the least upper bound of D_0. Then we can form

$$a_1 < f(a_1) < f^2(a_1) < \cdots$$

in the same way to obtain D_1, and we can continue this process, to obtain

$$D_1, D_2, \ldots.$$

It is clear that $D_1, D_2, \ldots$ are contained in M. If we had a precise way of expressing the fact that we can establish a never-ending string of such denumerable sets, then we would obtain what we want. The point is that we are now trying to prove Zorn's lemma, which is the natural tool for guaranteeing the existence of such a string. However, given such a string, we observe that its elements have two properties: If c is an element of such a string and $x < c$, then $f(x) \leqq c$. Furthermore, there is no element between c and $f(c)$, that is if x is an element of the string, then $x \leqq c$ or $f(c) \leqq x$. We shall now prove two lemmas which show that elements of M have these properties.]

Let $c \in M$. We shall say that c is an **extreme point** of M if whenever $x \in M$ and $x < c$, then $f(x) \leqq c$. For each extreme point $c \in M$ we let

$$M_c = \text{set of } x \in M \quad \text{such that} \quad x \leqq c \quad \text{or} \quad f(c) \leqq x.$$

Note that M_c is not empty because a is in it.

Lemma 2.2. *We have $M_c = M$ for every extreme point c of M.*

Proof. It will suffice to prove that M_c is an admissible subset. Let $x \in M_c$. If $x < c$ then $f(x) \leqq c$ so $f(x) \in M_c$. If $x = c$ then $f(x) = f(c)$ is again in M_c. If $f(c) \leqq x$, then $f(c) \leqq x \leqq f(x)$, so once more $f(x) \in M_c$. Thus we have proved that $f(M_c) \subset M_c$.

Let T be a totally ordered subset of M_c and let b be the least upper bound of T in M. If all elements $x \in T$ are $\leqq c$, then $b \leqq c$ and $b \in M_c$. If some $x \in T$ is such that $f(c) \leqq x$, then $f(c) \leqq x \leqq b$, and so b is in M_c. This proves our lemma.

Lemma 2.3. *Every element of M is an extreme point.*

Proof. Let E be the set of extreme points of M. Then E is not empty because $a \in E$. It will suffice to prove that E is an admissible subset. We first prove that f maps E into itself. Let $c \in E$. Let $x \in M$ and suppose $x < f(c)$. We must prove that $f(x) \leqq f(c)$. By Lemma 2.2, $M = M_c$, and hence we have $x < c$, or $x = c$, or $f(c) \leqq x$. This last possibility cannot occur because $x < f(c)$. If $x < c$ then

$$f(x) \leqq c \leqq f(c).$$

If $x = c$ then $f(x) = f(c)$, and hence $f(E) \subset E$.

Next let T be a totally ordered subset of E. Let b be the least upper bound of T in M. We must prove that $b \in E$. Let $x \in M$ and $x < b$. If for all $c \in T$ we have $f(c) \leqq x$, then $c \leqq f(c) \leqq x$ implies that x is an upper bound for T, whence $b \leqq x$, which is impossible. Since $M_c = M$ for all $c \in E$, we must therefore have $x \leqq c$ for some $c \in T$. If $x < c$, then $f(x) \leqq c \leqq b$, and if $x = c$, then

$$c = x < b.$$

Since c is an extreme point and $M_c = M$, we get $f(x) \leqq b$. This proves that $b \in E$, that E is admissible, and thus proves Lemma 2.3.

We now see trivially that M is totally ordered. For let $x, y \in M$. Then x is an extreme point of M by Lemma 2, and $y \in M_x$ so $y \leqq x$ or

$$x \leqq f(x) \leqq y,$$

thereby proving that M is totally ordered. As remarked previously, this concludes the proof of Theorem 2.1.

We shall obtain Zorn's lemma essentially as a corollary of Theorem 2.1. We first obtain Zorn's lemma in a slightly weaker form.

Corollary 2.4. *Let A be a non-empty strictly inductively ordered set. Then A has a maximal element.*

Proof. Suppose that A does not have a maximal element. Then for each $x \in A$ there exists an element $y_x \in A$ such that $x < y_x$. Let $f: A \to A$ be the map such that $f(x) = y_x$ for all $x \in A$. Then A, f satisfy the hypotheses of Theorem 2.1 and applying Theorem 2.1 yields a contradiction.

The only difference between Corollary 2.4 and Zorn's lemma is that in Corollary 2.4, we assume that a non-empty totally ordered subset has a *least* upper bound, rather than an upper bound. It is, however, a simple matter to reduce Zorn's lemma to the seemingly weaker form of Corollary 2.4. We do this in the second corollary.

Corollary 2.5. (Zorn's lemma). *Let S be a non-empty inductively ordered set. Then S has a maximal element.*

Proof. Let A be the set of non-empty totally ordered subsets of S. Then A is not empty since any subset of S with one element belongs to A. If $X, Y \in A$, we define $X \leqq Y$ to mean $X \subset Y$. Then A is partially ordered, and is in fact strictly inductively ordered. For let $T = \{X_i\}_{i \in I}$ be a totally ordered subset of A. Let

$$Z = \bigcup_{i \in I} X_i.$$

Then Z is totally ordered. To see this, let $x, y \in Z$. Then $x \in X_i$ and $y \in X_j$ for some $i, j \in I$. Since T is totally ordered, say $X_i \subset X_j$. Then $x, y \in X_j$ and since X_j is totally ordered, $x \leqq y$ or $y \leqq x$. Thus Z is totally ordered, and is obviously a least upper bound for T in A. By Corollary 2.4, we conclude that A has a maximal element X_0. This means that X_0 is a maximal totally ordered subset of S (non-empty). Let m be an upper bound for X_0 in S. Then m is the desired maximal element of S. For if $x \in S$ and $m \leqq x$ then $X_0 \cup \{x\}$ is totally ordered, whence equal to X_0 by the maximality of X_0. Thus $x \in X_0$ and $x \leqq m$. Hence $x = m$, as was to be shown.

§3. CARDINAL NUMBERS

Let A, B be sets. We shall say that the **cardinality** *of A is the same as the cardinality of B*, and write

$$\operatorname{card}(A) = \operatorname{card}(B)$$

if there exists a bijection of A onto B.

We say card(A) $\leqq$ card(B) if there exists an injective mapping (injection) $f: A \to B$. We also write card(B) $\geqq$ card(A) in this case. It is clear that if card(A) $\leqq$ card(B) and card(B) $\leqq$ card(C), then card(A) $\leqq$ card(C).

This amounts to saying that a composite of injective mappings is injective. Similarly, if card(A) = card(B) and card(B) = card(C) then card(A) = card(C).

This amounts to saying that a composite of bijective mappings is bijective. We clearly have card(A) = card(A). Using Zorn's lemma, it is easy to show (see Exercise 14) that

$$\text{card}(A \leqq \text{card}(B) \quad \text{or} \quad \text{card}(B) \leqq \text{card}(A).$$

Let $f: A \to B$ be a surjective map of a set A onto a set B. Then

$$\text{card}(B) \leqq \text{card}(A).$$

This is easily seen, because for each $y \in B$ there exists an element $x \in A$, denoted by x_y, such that $f(x_y) = y$. Then the association $y \mapsto x_y$ is an injective mapping of B into A, whence by definition, card(B) $\leqq$ card(A).

Given two nonempty sets A, B we have card(A) $\leqq$ card(B) *or* card(B) $\leqq$ card(A).

This is a simple application of Zorn's lemma. We consider the family of pairs (S, f) where S is a subset of A and $f: S \to B$ is an injective mapping. From the existence of a maximal element, the assertion follows at once.

Theorem 3.1. (Schroeder-Bernstein). *Let A, B be sets, and suppose that* card(A) $\leqq$ card(B), *and* card(B) $\leqq$ card(A). *Then*

$$\text{card}(A) = \text{card}(B).$$

Proof. Let

$$f: A \to B \quad \text{and} \quad g: B \to A$$

be injections. We separate A into two disjoint sets A_1 and A_2. We let A_1 consist of all $x \in A$ such that, when we lift back x by a succession of inverse maps,

$$x, g^{-1}(x), \quad f^{-1} \circ g^{-1}(x), \quad g^{-1} \circ f^{-1} \circ g^{-1}(x), \ldots$$

then at some stage we reach an element of A which cannot be lifted back to B by g. We let A_2 be the complement of A_1, in other words, the set of $x \in A$ which can be lifted back indefinitely, or such that we get stopped in B (i.e. reach an element of B which has no inverse image in A by f). Then $A = A_1 \cup A_2$. We shall define a bijection h of A onto B.

If $x \in A_1$, we define $h(x) = f(x)$.

If $x \in A_2$, we define $h(x) = g^{-1}(x) =$ unique element $y \in B$ such that $g(y) = x$.

Then trivially, h is injective. We must prove that h is surjective. Let $b \in B$. If, when we try to lift back b by a succession of maps

$$\cdots \circ f^{-1} \circ g^{-1} \circ f^{-1} \circ g^{-1} \circ f^{-1}(b)$$

we can lift back indefinitely, or if we get stopped in B, then $g(b)$ belongs to A_2 and consequently $b = h(g(b))$, so b lies in the image of h. On the other hand, if we cannot lift back b indefinitely, and get stopped in A, then $f^{-1}(b)$ is defined (i.e., b is in the image of f), and $f^{-1}(b)$ lies in A_1. In this case, $b = H(f^{-1}(b))$ is also in the image of h, as was to be shown.

Next we consider theorems concerning sums and products of cardinalities. We shall reduce the study of cardinalities of products of arbitrary sets to the denumerable case, using Zorn's lemma. Note first that an infinite set A always contains a denumerable set. Indeed, since A is infinite, we can first select an element $a_1 \in A$, and the complement of $\{a_1\}$ is infinite. Inductively, if we have selected distinct elements $a_1, \ldots, a_n$ in A, the complement of $\{a_1, \ldots, a_n\}$ is infinite, and we can select a_{n+1} in this complement. In this way, we obtain a sequence of distinct elements of A, giving rise to a denumerable subset of A.

Let A be a set. By a **covering** of A one means a set Γ of subsets of A such that the union

$$\bigcup_{C \in \Gamma} C$$

of all the elements of Γ is equal to A. We shall say that Γ is a **disjoint covering** if whenever $C, C' \in \Gamma$, and $C \neq C'$, then the intersection of C and C' is empty.

Lemma 3.2. *Let A be an infinite set. Then there exists a disjoint covering of A by denumerable sets.*

Proof. Let S be the set whose elements are pairs (B, Γ) consisting of a subset B of A, and a disjoint covering of B by denumerable sets. Then S is not empty. Indeed, since A is infinite, A contains a denumerable set D, and the pair $(D, \{D\})$ is in S. If (B, Γ) and (B', Γ') are elements of S, we define

$$(B, \Gamma) \leqq (B', \Gamma')$$

to mean that $B \subset B'$, and $\Gamma \subset \Gamma'$. Let T be a totally ordered non-empty subset of S. We may write $T = \{(B_i, \Gamma_i)\}_{i \in I}$ for some indexing set I. Let

$$B = \bigcup_{i \in I} B_i \quad \text{and} \quad \Gamma = \bigcup_{i \in I} \Gamma_i.$$

If $C, C' \in \Gamma$, $C \neq C'$, then there exists some indices i, j such that $C \in \Gamma_i$ and $C' \in \Gamma_j$. Since T is totally ordered, we have, say,

$$(B_i, \Gamma_i) \leqq (B_j, \Gamma_j).$$

Hence in fact, C, C' are both elements of Γ_j, and hence C, C' have an empty intersection. On the other hand, if $x \in B$, then $x \in B_i$ for some i, and hence there is some $C \in \Gamma_i$ such that $x \in C$. Hence Γ is a disjoint covering of B. Since the

elements of each Γ_i are denumerable subsets of A, it follows that Γ is a disjoint covering of B by denumerable sets, so (B, Γ) is in S, and is obviously an upper bound for T. Therefore S is inductively ordered.

Let (M, Δ) be a maximal element of S, by Zorn's lemma. Suppose that $M \neq A$. If the complement of M in A is infinite, then there exists a denumerable set D contained in this complement. Then

$$(M \cup D, \Delta \cup \{D\})$$

is a bigger pair than (M, Δ), contradicting the maximality of (M, Δ). Hence the complement of M in A is a finite set F. Let D_0 be an element of Δ. Let

$$D_1 = D_0 \cup F.$$

Then D_1 is denumerable. Let Δ_1 be the set consisting of all elements of Δ, except D_0, together with D_1. Then Δ_1 is a disjoint covering of A by denumerable sets, as was to be shown.

Theorem 3.3. *Let A be an infinite set, and let D be a denumerable set. Then*

$$\mathrm{card}(A \times D) = \mathrm{card}(A).$$

Proof. By the lemma, we can write

$$A = \bigcup_{i \in I} D_i$$

as a disjoint union of denumerable sets. Then

$$A \times D = \bigcup_{i \in I} (D_i \times D).$$

For each $i \in I$, there is a bijection of $D_i \times D$ on D_i by Proposition 1.5. Since the sets $D_i \times D$ are disjoint, we get in this way a bijection of $A \times D$ on A, as desired.

Corollary 3.4. *If F is a finite non-empty set, then*

$$\mathrm{card}(A \times F) = \mathrm{card}(A).$$

Proof. We have

$$\mathrm{card}(A) \leqq \mathrm{card}(A \times F) \leqq \mathrm{card}(A \times D) = \mathrm{card}(A).$$

We can then use Theorem 3.1 to get what we want.

Corollary 3.5. *Let A, B be non-empty sets, A infinite, and suppose*

$$\mathrm{card}(B) \leqq \mathrm{card}(A).$$

Then

$$\text{card}(A \cup B) = \text{card}(A).$$

Proof. We can write $A \cup B = A \cup C$ for some subset C of B, such that C and A are disjoint. (We let C be the set of all elements of B which are not elements of A.) Then $\text{card}(C) \leqq \text{card}(A)$. We can then construct an injection of $A \cup C$ into the product

$$A \times \{1, 2\}$$

of A with a set consisting of 2 elements. Namely, we have a bijection of A with $A \times \{1\}$ in the obvious way, and also an injection of C into $A \times \{2\}$. Thus

$$\text{card}(A \cup C) \leqq \text{card}(A \times \{1, 2\}).$$

We conclude the proof by Corollary 3.4 and Theorem 3.1.

Theorem 3.6. *Let A be an infinite set. Then*

$$\text{card}(A \times A) = \text{card}(A).$$

Proof. Let S be the set consisting of pairs (B, f) where B is an infinite subset of A, and $f : B \times B$ is a bijection of B onto $B \times B$. Then S is not empty because if D is a denumerable subset of A, we can always find a bijection of D on $D \times D$. If (B, f) and (B', f') are in S, we define $(B, f) \leqq (B', f')$ to mean $B \subset B'$, and the restriction of f' to B is equal to f. Then S is partially ordered, and we contend that S is inductively ordered. Let T be a non-empty totally ordered subset of S, and say T consists of the pairs (B_i, f_i) for i in some indexing set I. Let

$$M = \bigcup_{i \in I} B_i.$$

We shall define a bijection $g : M \to M \times M$. If $x \in M$, then x lies in some B_i. We define $g(x) = f_i(x)$. This value $f_i(x)$ is independent of the choice of B_i in which x lies. Indeed, if $x \in B_j$ for some $j \in I$, then say

$$(B_i, f_i) \leqq (B_j, f_j).$$

By assumption, $B_i \subset B_j$, and $f_j(x) = f_i(x)$, so g is well defined. To show g is surjective, let $x, y \in M$ and $(x, y) \in M \times M$. Then $x \in B_i$ for some $i \in I$ and $y \in B_j$ for some $j \in I$. Again since T is totally ordered, say $(B_i, f_i) \leqq (B_j, f_j)$. Thus $B_i \subset B_j$, and $x, y \in B_j$. There exists an element $b \in B_j$ such that

$$f_j(b) = (x, y) \in B_j \times B_j.$$

By definition, $g(b) = (x, y)$; so g is surjective. We leave the proof that g is injective to the reader to conclude the proof that g is a bijection. We then see

that (M, g) is an upper bound for T in S, and therefore that S is inductively ordered.

Let (M, g) be a maximal element of S, and let C be the complement of M in A. If $\operatorname{card}(C) \leqq \operatorname{card}(M)$, then

$$\operatorname{card}(M) \leqq \operatorname{card}(A) = \operatorname{card}(M \cup C) = \operatorname{card}(M)$$

by Corollary 3.5, and hence $\operatorname{card}(M) = \operatorname{card}(A)$ by Bernstein's Theorem. Since $\operatorname{card}(M) = \operatorname{card}(M \times M)$, we are done with the proof in this case. If

$$\operatorname{card}(M) \leqq \operatorname{card}(C),$$

then there exists a subset M_1 of C having the same cardinality as M. We consider

$$\begin{aligned}(M \cup M_1) \times (M \cup M_1)& \\ &= (M \times M) \cup (M_1 \times M) \cup (M \times M_1) \cup (M_1 \times M_1).\end{aligned}$$

By the assumption on M and Corollary 3.5, the last three sets in parentheses on the right of this equation have the same cardinality as M. Thus

$$(M \cup M_1) \times (M \cup M_1) = (M \times M) \cup M_2$$

where M_2 is disjoint from $M \times M$, and has the same cardinality as M. We now define a bijection

$$g_1 : M \cup M_1 \to (M \cup M_1) \times (M \cup M_1).$$

We let $g_1(x) = g(x)$ if $x \in M$, and we let g_1 on M_1 be any bijection of M_1 on M_2. In this way we have extended g to $M \cup M_1$, and the pair $(M \cup M_1, g_1)$ is in S, contradicting the maximality of (M, g). The case $\operatorname{card}(M) \leqq \operatorname{card}(C)$ therefore cannot occur, and our theorem is proved (using Exercise 14 below).

Corollary 3.7. *If A is an infinite set, and $A^{(n)} = A \times \cdots \times A$ is the product taken n times, then*

$$\operatorname{card}(A^{(n)}) = \operatorname{card}(A).$$

Proof. Induction.

Corollary 3.8. *If $A_1, \ldots, A_n$ are non-empty sets with A_n infinite, and*

$$\operatorname{card}(A_i) \leqq \operatorname{card}(A_n)$$

for $i = 1, \ldots, n$, then

$$\operatorname{card}(A_1 \times \cdots \times A_n) = \operatorname{card}(A_n).$$

Proof. We have

$$\text{card}(A_n) \leqq \text{card}(A_1 \times \cdots \times A_n) \leqq \text{card}(A_n \times \cdots \times A_n)$$

and we use Corollary 3.7 and the Schroeder-Bernstein theorem to conclude the proof.

Corollary 3.9. *Let A be an infinite set, and let Φ be the set of finite subsets of A. Then*

$$\text{card}(\Phi) = \text{card}(A).$$

Proof. Let Φ_n be the set of subsets of A having exactly n elements, for each integer $n = 1, 2, \ldots$. We first show that $\text{card}(\Phi_n) \leqq \text{card}(A)$. If F is an element of Φ_n, we order the elements of F in any way, say

$$F = \{x_1, \ldots, x_n\}.$$

and we associate with F the element $(x_1, \ldots, x_n) \in A^{(n)}$,

$$F \mapsto (x_1, \ldots, x_n).$$

If G is another subset of A having n elements, say $G = \{y_1, \ldots, y_n\}$, and $G \neq F$, then

$$(x_1, \ldots, x_n) \neq (y_1, \ldots, y_n).$$

Hence our map

$$F \mapsto (x_1, \ldots, x_n)$$

of Φ_n into $A^{(n)}$ is injective. By Corollary 3.7, we conclude that

$$\text{card}(\Phi_n) \leqq \text{card}(A).$$

Now Φ is the disjoint union of the Φ_n for $n = 1, 2, \ldots$ and it is an exercise to show that $\text{card}(\Phi) \leqq \text{card}(A)$ (cf. Exercise 1). Since

$$\text{card}(A) \leqq \text{card}(\Phi),$$

because in particular, $\text{card}(\Phi_1) = \text{card}(A)$, we see that our corollary is proved.

In the next theorem, we shall see that given a set, there always exists another set whose cardinality is bigger.

Theorem 3.10. *Let A be an infinite set, and T the set consisting of two elements $\{0, 1\}$. Let M be the set of all maps of A into T. Then*

$$\text{card}(A) \leqq \text{card}(M) \quad \textit{and} \quad \text{card}(A) \neq \text{card}(M).$$

Proof. For each $x \in A$ we let

$$f_x : A \to \{0, 1\}$$

be the map such that $f_x(x) = 1$ and $f_x(y) = 0$ if $y \neq x$. Then $x \mapsto f_x$ is obviously an injection of A into M, so that $\text{card}(A) \leqq \text{card}(M)$. Suppose that

$$\text{card}(A) = \text{card}(M).$$

Let

$$x \mapsto g_x$$

be a bijection between A and M. We define a map $h : A \to \{0, 1\}$ by the rule

$$h(x) = 0 \quad \text{if} \quad g_x(x) = 1,$$

$$h(x) = 1 \quad \text{if} \quad g_x(x) = 0.$$

Then certainly $h \neq g_x$ for any x, and this contradicts the assumption that $x \mapsto g_x$ is a bijection, thereby proving Theorem 3.10.

Corollary 3.11. *Let A be an infinite set, and let S be the set of all subsets of A. Then* $\text{card}(A) \leqq \text{card}(S)$ *and* $\text{card}(A) \neq \text{card}(S)$.

Proof. We leave it as an exercise. [*Hint*: If B is a non-empty subset of A, use the characteristic function φ_B such that

$$\varphi_B(x) = 1 \quad \text{if} \quad x \in B,$$

$$\varphi_B(x) = 0 \quad \text{if} \quad x \notin B.$$

What can you say about the association $B \mapsto \varphi_B$?]

§4. WELL-ORDERING

An ordered set A is said to be **well-ordered** if it is totally ordered, and if every non-empty subset B has a least element, that is, an element $a \in B$ such that $a \leqq x$ for all $x \in B$.

Example 1. The set of positive integers $\mathbf{Z}^+$ is well-ordered. Any finite set can be well-ordered, and a denumerable set D can be well-ordered: Any bijection of D with $\mathbf{Z}^+$ will give rise to a well-ordering of D.

Example 2. Let S be a well-ordered set and let b be an element of some set, $b \notin S$. Let $A = S \cup \{b\}$. We define $x \leqq b$ for all $x \in S$. Then A is totally ordered, and is in fact well-ordered.

Proof. Let B be a non-empty subset of A. If B consists of b alone, then b is a least element of B. Otherwise, B contains some element $a \in A$. Then $B \cap A$ is not empty, and hence has a least element, which is obviously also a least element for B.

Theorem 4.1. *Every non-empty set can be well-ordered.*

Proof. Let A be a non-empty set. Let S be the set of all pairs (X, ω), where X is a subset of A and ω is a well-ordering of X. Note that S is not empty because any single element of A gives rise to such a pair. If (X, ω) and (X', ω') are such pairs, we define $(X, \omega) \leqq (X', \omega')$ if $X \subset X'$, if the ordering induced on X by ω' is equal to ω, and if X is the initial segment of X'. It is obvious that this defines an ordering on S, and we contend that S is inductively ordered. Let $\{(X_i, \omega_i)\}$ be a totally ordered non-empty subset of S. Let $X = \bigcup X_i$. If $a, b \in X$, then a, b lie in some X_i, and we define $a \leqq b$ in X if $a \leqq b$ with respect to the ordering ω_i. This is independent of the choice of i (immediate from the assumption of total ordering). In fact, X is well ordered, for if Y is a non-empty subset of X, then there is some element $y \in Y$ which lies in some X_j. Let c be a least element of $X_j \cap Y$. One verifies at once that c is a least element of Y. We can therefore apply Zorn's lemma. Let (X, ω) be a maximal element in S. If $X \neq A$, then, using Example 2, we can define a well-ordering on a bigger subset than X, contradicting the maximality assumption. This proves Theorem 4.1.

Note. Theorem 4.1 is an immediate and straightforward consequence of Zorn's lemma. Usually in mathematics, Zorn's lemma is the most efficient tool when dealing with infinite processes.

EXERCISES

1. Prove the statement made in the proof of Corollary 3.9.
2. If A is an infinite set, and Φ_n is the set of subsets of A having exactly n elements, show that

$$\operatorname{card}(A) \leqq \operatorname{card}(\Phi_n)$$

for $n \geqq 1$.

3. Let A_i be infinite sets for $i = 1, 2, \ldots$ and assume that

$$\operatorname{card}(A_i) \leqq \operatorname{card}(A)$$

for some set A, and all i. Show that

$$\operatorname{card}\left(\bigcup_{i=1}^{\infty} A_i\right) \leqq \operatorname{card}(A).$$

4. Let K be a subfield of the complex numbers. Show that for each integer $n \geqq 1$, the cardinality of the set of extensions of K of degree n in $\mathbf{C}$ is $\leqq \operatorname{card}(K)$.
5. Let K be an infinite field, and E an algebraic extension of K. Show that

$$\operatorname{card}(E) = \operatorname{card}(K).$$

6. Finish the proof of the Corollary 3.11.
7. If A, B are sets, denote by $M(A, B)$ the set of all maps of A into B. If B, B' are sets with the same cardinality, show that $M(A, B)$ and $M(A, B')$ have the same cardinality. If A, A' have the same cardinality, show that $M(A, B)$ and $M(A', B)$ have the same cardinality.
8. Let A be an infinite set and abbreviate $\operatorname{card}(A)$ by α. If B is an infinite set, abbreviate $\operatorname{card}(B)$ by β. Define $\alpha\beta$ to be $\operatorname{card}(A \times B)$. Let B' be a set disjoint from A such that $\operatorname{card}(B) = \operatorname{card}(B')$. Define $\alpha + \beta$ to be $\operatorname{card}(A \cup B')$. Denote by B^A the set of all maps of A into B, and denote $\operatorname{card}(B^A)$ by β^α. Let C be an infinite set and abbreviate $\operatorname{card}(C)$ by γ. Prove the following statements:
   (a) $\alpha(\beta + \gamma) = \alpha\beta + \alpha\gamma$.
   (b) $\alpha\beta = \beta\alpha$.
   (c) $\alpha^{\beta+\gamma} = \alpha^\beta\alpha^\gamma$.
9. Let K be an infinite field. Prove that there exists an algebraically closed field K^a containing K as a subfield, and algebraic over K. [*Hint*: Let Ω be a set of cardinality strictly greater than the cardinality of K, and containing K. Consider the set S of all pairs (E, φ) where E is a subset of Ω such that $K \subset E$, and φ denotes a law of addition and multiplication on E which makes E into a field such that K is a subfield, and E is algebraic over K. Define a partial ordering on S in an obvious way; show that S is inductively ordered, and that a maximal element is algebraic over K and algebraically closed. You will need Exercise 5 in the last step.]
10. Let K be an infinite field. Show that the field of rational functions $K(t)$ has the same cardinality as K.
11. Let J_n be the set of integers $\{1, \ldots, n\}$. Let $\mathbf{Z}^+$ be the set of positive integers. Show that the following sets have the same cardinality:
   (a) The set of all maps $M(\mathbf{Z}^+, J_n)$.
   (b) The set of all maps $M(\mathbf{Z}^+, J_2)$.
   (c) The set of all real numbers x such that $0 \leqq x < 1$.
   (d) The set of all real numbers.
12. Show that $M(\mathbf{Z}^+, \mathbf{Z}^+)$ has the same cardinality as the real numbers.
13. Let S be a non-empty set. Let S' denote the product S with itself taken denumerably many times. Prove that $(S')'$ has the same cardinality as S'. [Given a set S whose cardinality is strictly greater than the cardinality of $\mathbf{R}$, I do not know whether it is always true that card S = card S'.] Added 1994: The grapevine communicates to me that according to Solovay, the answer is "no."
14. Let A, B be non–empty sets. Prove that

$$\operatorname{card}(A) \leqq \operatorname{card}(B) \quad \text{or} \quad \operatorname{card}(B) \leqq \operatorname{card}(A).$$

[Hint: consider the family of pairs (C, f) where C is a subset of A and $f\colon C \to B$ is an injective map. By Zorn's lemma there is a maximal element. Now finish the proof].

Bibliography

[Ad 62] F. ADAMS, Vector Fields on Spheres, *Ann. Math.* **75** (1962) pp. 603–632

[Ara 31] H. ARAMATA, Über die Teilbarkeit der Dedekindschen Zetafunktionen, *Proc. Imp. Acad. Tokyo* **7** (1931) pp. 334–336

[Ara 33] H. ARAMATA, Über die Teilbarkeit der Dedekindschen Zetafunktionen, *Proc. Imp. Acad. Tokyo* **9** (1933) pp. 31–34

[Art 24] E. ARTIN, Kennzeichnumg des Körpers der reellen algebraischen Zahlen, *Abh. Math. Sem. Hansischen Univ.* **3** (1924) pp. 319–323

[Art 27] E. ARTIN, Über die Zerlegung definiter Funktionen in Quadrate, *Abh. Math. Sem. Hansischen Univ.* **5** (1927) pp. 100–115

[Art 44] E. ARTIN, *Galois Theory*, University of Notre Dame, 1944

[ArS 27] E. ARTIN and E. SCHREIER, Algebraische Konstruktion reeller Körper, *Abh. Math. Sem. Hansischen Univ.* **5** (1927) pp. 85–99

[ArT 68] E. ARTIN and J. TATE, *Class Field Theory*, Benjamin-Addison Wesley, 1968 (reprinted by Addison-Wesley, 1991)

[Art 68] M. ARTIN, On the solutions of analytic equations, *Invent. Math.* **5** (1968) pp. 277–291

[ArM 65] M. ARTIN and B. MAZUR, On periodic points, *Ann. Math. (2)* **81** (1965) pp. 89–99

[At 61] M. ATIYAH, Characters and cohomology of finite groups, *Pub. IHES* **9** (1961) pp. 5–26

[At 67] M. ATIYAH, *K-Theory*, Addison-Wesley, (reprinted from the Benjamin Lecture Notes, 1967)

[ABP 73] M. ATIYAH, R. BOTT, V. PATODI, On the heat equation and the index theorem, *Invent. Math.* **19** (1973) pp. 270–330

[ABS 64] M. ATIYAH, R. BOTT, A. SHAPIRO, Clifford Modules, *Topology* **Vol. 3 Supp. 1** (1964) pp. 3–38

[AtM 69] M. ATIYAH and I. MCDONALD, *Introduction to commutative algebra*, Addison-Wesley, 1969

[Ba 68] H. BASS, *Algebraic K-theory*, Benjamin, 1968

[BaH 62] P. T. BATEMAN and R. HORN, A heuristic asymptotic formula concerning the distribution of prime numbers, *Math. Comp.* **16** (1962) pp. 363–367

[Be 80] G. BELYI, Galois extensions of the maximal cyclotomic field, *Izv. Akad. Nauk SSSR* **43** (1979) pp. 267–276 (= *Math. USSR Izv.* **14** (1980), pp. 247–256

[Be 83] G. BELYI, On extensions of the maximal cyclotomic field having a given classical Galois group, *J. reine angew. Math.* **341** (1983) pp. 147–156

[BeY 91] C. BERENSTEIN and A. YGER, Effective Bezout identities in Q$[z_1, \ldots z_n]$, *Acta Math.* **166** (1991) pp. 69–120

[BGV 92] N. BERLINE, E. GETZLER, M. VERGNE, *Heat kernels and Dirac operators*, Springer-Verlag, 1992

[BCHS 65] B. BIRCH, S. CHOWLA, M. HALL, and A. SCHINZEL, On the difference $x^3 - y^2$, *Norske Vid. Selsk. Forrh.* **38** (1965) pp. 65–69

[Bott 69] R. BOTT, *Lectures on $K(X)$*, Benjamin 1969

[Boun 1854] V. BOUNIAKOWSKY, Sur les diviseurs numériques invariables des fonctions

rationnelles entières, *Mémoires sc. math. et phys.* **T. VI** (1854–1855) pp. 307–329

[Bour 82] N. BOURBAKI, *Lie algebras and Lie groups*, Masson, 1982

[Bra 47a] R. BRAUER, On the zeta functions of algebraic number fields, *Amer. J. Math.* **69** (1947) pp. 243–250

[Bra 47b] R. BRAUER, On Artin's L-series with general group characters, *Ann. Math.* **48** (1947) pp. 502–514

[BLSTW 83] J. BRILLHART, D. H. LEHMER, J. L. SELFRIDGE, B. TUCKERMAN, and S. WAGSTAFF, Factorization of $b^n \pm 1$, $b = 2, 3, 5, 6, 7, 10, 11$ up to high powers, *Contemporary Mathematics* **Vol. 22**, AMS, Providence, RI, 1983

[BtD 85] T. BRÖCKER and T. TOM DIECK, *Representations of Compact Lie Groups*, Springer-Verlag, 1985

[Br 87] D. BROWNAWELL, Bounds for the degree in Nullstellensatz, *Ann. of Math.* **126** (1987) pp. 577–592

[Br 88] D. BROWNAWELL, Local diophantine nullstellen inequalities, *J. Amer. Math. Soc.* **1** (1988) pp. 311–322

[Br 89] D. BROWNAWELL, Applications of Cayley-Chow forms, *Springer Lecture Notes* **1380**: *Number Theory, Ulm*, 1987, H. P. Schlickewei and E. Wirsing (eds.) pp. 1–18

[BrCDT 01] C. BREUIL, B. CONRAD, F. DIAMOND, R. TAYLOR, On the modularity of elliptic curves over **Q**: Wild 3-adic exercises, *J. Amer. Math. Soc.* **14** (2001) pp. 843–839

[CaE 57] E. CARTAN and S. EILENBERG, Homological Algebra, Princeton University Press, 1957

[CCFT 91] P. CASSOU-NOGUES, T. CHINBURG, A. FROLICH, M. J. TAYLOR, L-functions and Galois modules, in *L-functions and Arithmetic*, J. Coates and M. J. Taylor (eds.), Proceedings of the Durham symposium July 1989, *London Math. Soc. Lecture Notes Series* **153**, Cambridge University Press (1991) pp. 75–139

[CuR 81] C. W. CURTIS and I. REINER, *Methods of Representation Theory*, John Wiley and Sons, 1981

[Da 65] H. DAVENPORT, On $f^3(t) - g^2(t)$, *Norske Vid. Selsk. Forrh.* **38** (1965) pp. 86–87

[De 68] P. DELIGNE, Formes modulaires et représentations l-adiques, *Séminaire Bourbaki* 1968–1969, pp. 55–105

[De 73] P. DELIGNE, Formes modulaires et représentations de GL(2), *Springer Lecture Notes* **349** (1973) pp. 507–530

[DeS 74] P. DELIGNE and J.-P. SERRE, Formes modulaires de poids 1, *Ann. Sci. ENS* **7** (1974) pp. 507–530

[Dou 64] A. DOUADY, Determination d'un groupe de Galois, *C.R. Acad. Sci.* **258** (1964), pp. 5305–5308

[ES 52] S. EILENBERG and N. STEENROD, *Foundations of Algebraic Topology*, Princeton University Press, 1952

[Fa 91] G. FALTINGS, *Lectures on the arithmetic Riemann-Roch theorem*, Ann. Math. Studies **127**, 1991

[Fr 87] G. FREY, Links between stable elliptic curves and certain diophantine equations, *Number Theory, Lecture Notes* **1380**, Springer-Verlag 1987 pp. 31–62

[Fro 83] A. FRÖLICH, *Galois Module Stuctures of Algebraic Integers*, Ergebnisse der Math. 3 Folge Vol. 1, Springer-Verlag (1983)

[FuL 85] W. FULTON and S. LANG, Riemann-Roch Algebra, Springer-Verlag, 1985

[God 58] R. GODEMENT, *Théorie des faisceaux*, Hermann Paris, 1958

[Gor 68] D. GORENSTEIN, *Finite groups*, Harper and Row, 1968

[Gor 82] D. GORENSTEIN, *Finite simple groups*, Plenum Press, 1982

[Gor 83] D. GORENSTEIN, The classification of finite simple groups, Plenum Press, 1983

[Gor 86] D. GORENSTEIN, Classifying the finite simple groups, *Bull. AMS* **14** No. 1 (1986) pp. 1–98

[GreH 81] M. GREENBERG and J. HARPER, *Algebraic Topology: A First Course*, Benjamin-Addison Wesley, 1981

[GriH 78] P. GRIFFITHS and J. HARRIS, *Principles of Algebraic Geometry*, Wiley Interscience, New York, 1978

[Gro 57] A. GROTHENDIECK, Sur quelques points d'algèbre homologique, *Tohoku Math. J.* **9** (1957) pp. 119–221

[Gro 68] A. GROTHENDIECK, Classes de Chern et représentations linéaires des groupes discrets, *Dix exposés sur la cohomologie étale des schémas*, North-Holland, Amsterdam, 1968

[Gu 90] R. GUNNING, *Introduction to Holomorphic Functions of Several Variables*, Vol. II: Local Theory; Vol. III, Wadsworth and Brooks/Cole, 1990

[HalR 74] H. HALBERSTAM and H.-E. RICHERT, Sieve methods, Academic Press, 1974

[Hal 71] M. HALL, The diophantine equation $x^3 - y^2 = k$, *Computers and Number Theory*, ed. by A. O. L. Atkin and B. Birch, Academic Press, London 1971 pp. 173–198

[HardW 71] G. H. HARDY and E. M. WRIGHT, *An Introduction to the Theory of Numbers*, Oxford University Press, Oxford, UK, 1938–01971 (several editions)

[Hart 77] R. HARTSHORNE, *Algebraic Geometry*, Springer-Verlag, New York, 1977

[Has 34] H. HASSE, Abstrakte Begrundung der komplexen Multiplikation und Riemannsche Vermutung in Funktionenkörpern, *Abh. Math. Sem. Univ. Hamburg* **10** (1934) pp. 325–348

[HilS 70] P. J. HILTON and U. STAMMBACH, *A course in homological algebra*, Graduate Texts in Mathematics, Springer-Verlag 1970

[Hir 66] F. HIRZEBRUCH, *Topological methods in algebraic geometry*, Springer-Verlag, New York, 1966 (Translated and expanded from the original German, 1956)

[Hu 75] D. HUSEMOLLER, *Fibre Bundles*, Springer-Verlag, second edition, 1975

[Ih 66] Y. IHARA, On discrete subgroups of the two by two projective linear group over p-adic fields, *J. Math Soc. Japan* **18** (1966) pp. 219–235

[Ik 77] M. IKEDA, Completeness of the absolute Galois group of the rational number field, *J. reine angew. Math.* **291** (1977) pp. 1–22

[Iw 53] K. IWASAWA, On solvable extensions of algebraic number fields, *Ann. Math.* **548** (1953) pp. 548–572

[Ja 79] N. JACOBSON, *Lie algebras*, Dover, 1979 (reprinted from Interscience, 1962)

[Ja 85] N. JACOBSON, *Basic Algebra* I and II, second edition, Freeman, 1985

[JoL 01] J. JORGENSON and S. LANG, *Spherical Inversion on* $SL_n(\mathbf{R})$, Springer Verlag 2001

[Jou 80] J.-P. JOUANOLOU, Idéaux résultants, *Advances in Mathematics* **37 No. 3** (1980) pp. 212–238

[Jou 90] J.-P. JOUANOLOU, Le formalisme du résultant, *Advances in Mathematics* **90 No. 2** (1991) pp. 117–263

[Jou 91] J.-P. JOUANOLOU, *Aspects invariants de l'élimination*, Département de Mathématiques, Université Louis Pasteur, Strasbourg, France (1991)

[Ko 88] J. KOLLAR, Sharp effective nullstellensatz, *J. Amer. Math. Soc.* **1 No. 4** (1988) pp. 963–975

[Kr 32] W. KRULL, Allgemeine Bewertungstheorie, *J. reine angew. Math.* (1932) pp. 169–196

[La 52] S. LANG, On quasi algebraic closure, *Ann. Math.* **55** (1952) pp. 373–390

[La 53] S. LANG, The theory of real places, *Ann. Math.* **57** No. 2 (1953) pp. 378–391

[La 58] S. LANG, *Introduction to Algebraic Geometry*, Interscience, 1958

[La 70] S. LANG, *Algebraic Number Theory*, Addison-Wesley, 1970; reprinted by Springer-Verlag; second edition 1994

[La 72] S. LANG, *Differential Manifolds*, Addison-Wesley, 1972; reprinted by Springer-Verlag, 1985; superceded by [La 99a]

[La 73] S. LANG, *Elliptic Functions*, Springer-Verlag, 1973; second edition 1987

[La 76] S. LANG, *Introduction to Modular Forms*, Springer-Verlag 1976

[La 82] S. LANG, Units and class groups in number theory and algebraic geometry, *Bull. AMS* **Vol. 6 No. 3** (1982) pp. 253–316

[La 83] S. LANG, *Fundamentals of Diophantine Geometry*, Springer-Verlag 1983

[La 85] S. LANG, *Real Analysis*, Second edition, Addison-Wesley, 1985; third edition *Real and Functional Analysis*, Springer-Verlag, 1993

[La 90a] S. LANG, *Undergraduate Algebra*, second edition, Springer-Verlag, 1990

[La 90b] S. LANG, *Cyclotomic fields, I and II*, Springer-Verlag, New York, 1990, combined edition of the original editions, 1978, 1980

[La 90c] S. LANG, Old and new conjectured diophantine inequalities, *Bull. AMS* **Vol. 23 No. 1** (1990) pp. 37–75

[La 96] S. LANG, *Topics in Cohomology of Groups*, Springer Lecture Notes 1996, reproduced in Lang's *Collected Papers*, Vol. IV, Springer 2000

[La 99a] S. LANG, *Fundamentals of Differential Geometry*, Springer Verlag, 1999

[La 99b] S. LANG, *Math Talks for Undergraduates*, Springer Verlag, 1999

[LaT 75] S. LANG and H. TROTTER, *Distribution of Frobenius Elements in* GL_2*-Extensions of the Rational Numbers*, Springer Lecture Notes **504**

[Ma 16] F. MACAULAY, *The algebraic theory of modular systems*, Cambridge University Press, Cambridge UK, 1916

[Mack 51] G. MACKEY, On induced representations of groups, *Amer. J. Math.* **73** (1951) pp. 576–592

[Mack 53] G. MACKEY, Symmetric and anti-symmetric Kronecker squares of induced representations of finite groups, *Amer. J. Math.* **75** (1973) pp. 387–405

[Man 69] J. MANIN, *Lectures on the K-functor in algebraic geometry*, Russian Math. Surveys 24(5) (1969) pp. 1–89

[Man 71] A. MANNING, Axiom A diffeomorphisms have rational zeta functions, *Bull. Lond. Math. Soc.* **3** (1971) pp. 215–220

[Mas 84a] R. C. MASON, Equations over function fields, Springer Lecture Notes **1068** (1984) pp. 149–157; in *Number Theory, Proceedings of the Noordwijkerhout*, 1983

[Mas 84b] R. C. MASON, Diophantine equations over function fields, *London Math. Soc. Lecture Note Series* **Vol. 96**, Cambridge University Press, 1984

[Mas 84c] R. C. MASON, The hyperelliptic equation over function fields, *Math. Proc. Cambridge Philos. Soc.* **93** (1983) pp. 219–230

[MaW 85] D. MASSER and G. WÜSTHOLZ, Zero estimates on group varieties II, *Invent. Math.* **80** (1985) pp. 233–267

[Mat 80] H. MATSUMURA, *Commutative algebra*, second edition, Benjamin-Cummings, New York 1980

[Mat 86] H. MATSUMURA, *Commutative rings*, Cambridge University Press, 1986

[Matz 87] B. MATZAT, Konstruktive Galoistheorie, Springer Lecture Notes **1284**, 1987

[Matz 88] B. MATZAT, Über das Umkehrproblem der Galoischen Theorie, *Jahrsbericht Deutsch. Mat.-Verein.* **90** (1988) pp. 155–183

[Neu 69] J. NEUKIRCH, Über eine algebraische Kennzeichnung der Henselkörper, *J. reine angew. Math.* **231** (1968) pp. 75–81

[Neu 69] J. NEUKIRCH, Kennzeichnung der p-adische und endlichen algebraischen Zahlkörper, *Invent. Math.* **6** (1969) pp. 269–314

[Neu 69] J. NEUKIRCH, Kennzeichnung der endlich-algebraischen Zahlkörper durch die Galoisgruppe der maximal auflösbaren Erweiterungen, *J. fur Math.* **238** (1969) pp. 135–147

[No 76] D. NORTHCOTT, *Finite Free Resolutions*, Cambridge University Press, 1976

[Ph 86] P. PHILIPPON, Lemmes de zéros dans les groupes algébriques commutatifs, *Bull. Soc. Math. France* **114** (1986) pp. 355–383

[Ph 89] P. PHILIPPON, Sur des hauteurs alternatives I, *Math. Ann.* **289** (1991) pp. 255–283; II to appear

[Pop 94] F. POP, On birational anabelian geometry, *Annals of Math.* (1994) pp. 146–182

[Pop 95] F. POP, Etale Galois covers of affine smooth curves, *Invent. Math.* **120** (1995), pp. 555–578

[Ri 90a] K. RIBET, On modular representations of Gal($\mathbf{Q}^a/\mathbf{Q}$) arising from modular forms, *Invent. Math.* **100** (1990) pp. 431–476

[Rib 90b] K. RIBET, From the Taniyama-Shimura conjecture to Fermat's last theorem, *Annales de la Fac. des Sci. Toulouse* (1990) pp. 116–170

[Ric 60] C. RICKART, *Banach Algebras*, Van Nostrand (1960), Theorems 1.7.1 and 4.2.2

[Ro 79] J. ROTMAN, *Introduction to Homological Algebra*, Academic Press, 1979

[Ru 73] W. RUDIN, *Functional Analysis*, McGraw Hill (1973) Theorems 10.14, and 11.18

[Schi 58] A. SCHINZEL and W. SIERPINSKI, Sur certaines hypothèses concernant les nombres premiers, *Acta Arith.* **4** (1948) pp. 185–208

[Schulz 37] W. SCHULZ, Über die Galoissche Gruppe der Hermitschen Polynome, *J. reine angew. math.* **177** (1937) pp. 248–252

[Schur 31] J. SCHUR, Affektlose Gleichungen in der Theorie der Laguereschen und Hermiteschen Polynome, *J. reine angew. math.* **165** (1931) pp. 52–58

[Se 62] J.-P. SERRE, Endomorphismes complêtements continus des espaces de Banach p-adiques, *Pub. Math. IHES* **12** (1962) pp. 69–85

[Se 64] J.-P. SERRE, *Cohomologie Galoisienne*, Springer Lecture Notes **5**, 1964

[Se 65a] J.-P. SERRE, *Algèbre locale, multiplicités*, Springer Lecture Notes **11** (1965) Third Edition 1975

[Se 65b] J.-P. SERRE, *Lie algebras and Lie groups*, Benjamin, 1965; reprinted *Springer Lecture Notes* **1500**, Springer-Verlag 1992

[Se 68a] J.-P. SERRE, *Abelian l-adic representations and Elliptic Curves*, Benjamin, 1968

[Se 68b] J.-P. SERRE, Une interprétation des congruences relatives à la fonction de Ramanujan, *Séminaire Delange-Poitou-Pisot*, 1971–1972

[Se 72a] J.-P. SERRE, Propriétés Galoisiennes des points d'ordre fini des courbes elliptiques, *Invent. Math.* **15** (1972) pp. 259–331

[Se 72b] J.-P. SERRE, Congruences et formes modulaires (d'après Swinnerton-Dyer), *Séminaire Bourbaki*, 1971–1972

[Se 73] J.-P. SERRE, *A Course in Arithmetic*, Springer-Verlag, New York, 1973

[Se 80] J.-P. SERRE, *Trees*, Springer-Verlag, 1980

[Se 87] J.-P. SERRE, Sur les représentations modulaires de degré 2 de Gal($\mathbf{Q}^a/\mathbf{Q}$), *Duke Math. J.* **54** (1987), pp. 179–230

[Se 88] J.-P. SERRE, Groupes de Galois sur **Q**, *Séminaire Bourbaki*, 1987–1988, *Astérisque* **161–162**, pp. 73–85

[Se 92] J.-P. SERRE, *Topics in Galois theory*, course at Harvard, 1989, Jones and Bartlett, Boston 1992

[SGA 6] P. BERTHELOT, A. GROTHENDIECK, L. ILLUSIE et al., *Théorie des intersections et théorème de Riemann-Roch*, Springer Lecture Notes **146** (1967)

[Shaf 54] I. SHAFAREVICH, Construction of fields of algebraic number with given solvable Galois group, *Izv. Akad. Nauk SSSR* **18** (1954) pp. 525–578 (*Amer. math. Soc. Transl.* **4** (1956)) pp. 185–237

[Shat 72] S. SHATZ, *Profinite groups, arithmetic and geometry*, Ann. of Math. Studies, Princeton University Press 1972

[Shih 74] R.-Y. SHIH, On the construction of Galois extensions of function fields and number fields, *Math. Ann.* **207** (1974), pp. 99–120

[Shim 66] G. SHIMURA, A Reciprocity law in non-solvable extensions, *J. reine angew. Math.* **224** (1966) pp. 209–220

[Shim 71] G. SHIMURA, *Introduction to the arithmetic theory of Automorphic Functions*, Iwanami Shoten and Princeton University Press, 1971

[Shu 87] M. SHUB, *Global Stability of Dynamical Systems*, Springer-Verlag, New York 1987

[Sil 88] J. SILVERMAN, Wieferich's criterion and the *abc* conjecture, *J. Number Theory* **30** (1988) pp. 226–237

[Sn 00] N. SNYDER, An alternate proof of Mason's theorem, *Elemente der Math.* **55** (2000) pp. 93–94

[Sol 01] R. SOLOMON, A brief history of the classification of the finite simple groups, *Bull. AMS* Vol. **38** No. 3 (2001) pp. 315–322

[Sou 90] C. SOULÉ, Géometrie d'Arakelov et théorie des nombres transcendants, Preprint, 1990

[SteT 86] C.L. STEWART and R. TIJDEMAN, On the Oesterle–Masser Conjecture, *Mon. Math.* **102** (1986) pp. 251–257

[Sto 81] W. STOTHERS, Polynomial identites and hauptmoduln, *Quart. J. Math. Oxford* (2) **32** (1981) pp. 349–370

[Sw 69] R. SWAN, Invariant rational functions and a problem of Steenrod, *Invent. Math.* **7** (1969) pp. 148–158

[Sw 83] R. SWAN, Noether's problem in Galois theory, *Emmy Noether in Bryn Mawr*, J. D. Sally and B. Srinivasan, eds., Springer-Verlag, 1983, p. 40

[SwD 73] H. P. SWINNERTON-DYER, On l-adic representations and congruences for coefficients of modular forms, Antwerp conference, *Springer Lecture Notes* **350** (1973)

[TaW 95] R. TAYLOR and A. WILES, Ring-theoretic properties of certain Hecke algebras, *Annals of Math.* **141** (1995) pp. 553–572

[Uch 77] K. UCHIDA, Isomorphisms of Galois groups of algebraic function fields, *Ann. of Math.* **106** (1977) pp. 589–598

[Uch 79] K. UCHIDA, Isomorphisms of Galois groups of solvable closed Galois extensions, *Tohoku Math. J.* **31** (1979) pp. 359–362

[Uch 81] K. UCHIDA, Homomorphisms of Galois groups of solvably closed Galois extensions, *J. Math. Soc. Japan* **33** (1981) pp. 595–604

[vdW 29] B. L. VAN DER WAERDEN, On Hilbert's function, series of composition of ideals and a generalization of the theorem of Bezout, *Proc. R. Soc. Amsterdam* **31** (1929) pp. 749–770

[vdW 30] B. L. VAN DER WAERDEN, *Modern Algebra*, Springer-Verlag, 1930

[Wil 95] A. WILES, Modular elliptic curves and Fermat's last theorem, *Annals of Math.* **141** (1995) pp. 443–551

[Win 91] K. WINBERG, On Galois groups of p-closed algebraic number fields with restricted ramification, I, *J. reine angew. Math.* **400** (1989) pp. 185–202 and II, *ibid.* **416** (1991) pp. 187–194

[Wit 35] E. WITT, Der Existenzsatz für abelsche Funktionenkörper, *J. reine angew Math.* **173** (1936) pp. 43–51

[Wit 36] E. WITT, Konstruktion von galoisschen Körpern der der Charakteristic p mit vorgegebener Gruppe der Ordnung p^f, *J. reine angew. Math.* **174** (1936) pp. 237–245

[Wit 37] E. WITT, Zyklische Körper und Algebren der Charakteristik p vom Grad p^n. Struktur diskret bewerteter perfekter Körper mit vollkommenen Restklassenkörper der Charakteristik p, *J. reine angew. Math.* **176** (1937) pp. 126–140

[Za 95] U. ZANNIER, On Davenport's bound for the degree of $f^3 - g^2$ and Riemann's existence theorem, *Acta Arithm.* **LXXI.2** (1995) pp. 107–137

INDEX

Graduate Texts in Mathematics

1 TAKEUTI/ZARING. Introduction to Axiomatic Set Theory 2nd ed.
2 OXTOBY. Measure and Category. 2nd ed.
3 SCHAEFER. Topological Vector Spaces. 2nd ed.
4 HILTON/STAMMBACH A Course in Homological Algebra. 2nd ed.
5 MAC LANE. Categories for the Working Mathematician. 2nd ed.
6 HUGHES/PIPER. Projective Planes.
7 SERRE. A Course in Arithmetic.
8 TAKEUTI/ZARING. Axiomatic Set Theory.
9 HUMPHREYS. Introduction to Lie Algebras and Representation Theory.
10 COHEN A Course in Simple Homotopy Theory.
11 CONWAY. Functions of One Complex Variable I. 2nd ed.
12 BEALS. Advanced Mathematical Analysis.
13 ANDERSON/FULLER. Rings and Categories of Modules 2nd ed.
14 GOLUBITSKY/GUILLEMIN. Stable Mappings and Their Singularities.
15 BERBERIAN. Lectures in Functional Analysis and Operator Theory.
16 WINTER. The Structure of Fields.
17 ROSENBLATT. Random Processes. 2nd ed.
18 HALMOS. Measure Theory.
19 HALMOS. A Hilbert Space Problem Book. 2nd ed.
20 HUSEMOLLER. Fibre Bundles. 3rd ed.
21 HUMPHREYS. Linear Algebraic Groups.
22 BARNES/MACK. An Algebraic Introduction to Mathematical Logic.
23 GREUB. Linear Algebra. 4th ed.
24 HOLMES. Geometric Functional Analysis and Its Applications.
25 HEWITT/STROMBERG. Real and Abstract Analysis.
26 MANES. Algebraic Theories.
27 KELLEY. General Topology.
28 ZARISKI/SAMUEL. Commutative Algebra. Vol.I.
29 ZARISKI/SAMUEL. Commutative Algebra. Vol.II.
30 JACOBSON. Lectures in Abstract Algebra I. Basic Concepts.
31 JACOBSON. Lectures in Abstract Algebra II. Linear Algebra.
32 JACOBSON Lectures in Abstract Algebra III. Theory of Fields and Galois Theory.
33 HIRSCH. Differential Topology.
34 SPITZER. Principles of Random Walk. 2nd ed.
35 ALEXANDER/WERMER. Several Complex Variables and Banach Algebras. 3rd ed.
36 KELLEY/NAMIOKA et al. Linear Topological Spaces.
37 MONK. Mathematical Logic.
38 GRAUERT/FRITZSCHE. Several Complex Variables.
39 ARVESON. An Invitation to C^*-Algebras.
40 KEMENY/SNELL/KNAPP. Denumerable Markov Chains. 2nd ed.
41 APOSTOL. Modular Functions and Dirichlet Series in Number Theory 2nd ed.
42 SERRE Linear Representations of Finite Groups.
43 GILLMAN/JERISON. Rings of Continuous Functions.
44 KENDIG. Elementary Algebraic Geometry.
45 LOÈVE. Probability Theory I. 4th ed.
46 LOÈVE. Probability Theory II. 4th ed.
47 MOISE. Geometric Topology in Dimensions 2 and 3.
48 SACHS/WU. General Relativity for Mathematicians.
49 GRUENBERG/WEIR. Linear Geometry. 2nd ed.
50 EDWARDS. Fermat's Last Theorem.
51 KLINGENBERG A Course in Differential Geometry.
52 HARTSHORNE. Algebraic Geometry.
53 MANIN. A Course in Mathematical Logic.
54 GRAVER/WATKINS. Combinatorics with Emphasis on the Theory of Graphs.
55 BROWN/PEARCY. Introduction to Operator Theory I: Elements of Functional Analysis.
56 MASSEY. Algebraic Topology: An Introduction.
57 CROWELL/FOX. Introduction to Knot Theory.
58 KOBLITZ. p-adic Numbers, p-adic Analysis, and Zeta-Functions. 2nd ed.
59 LANG. Cyclotomic Fields.
60 ARNOLD. Mathematical Methods in Classical Mechanics. 2nd ed.
61 WHITEHEAD. Elements of Homotopy Theory.
62 KARGAPOLOV/MERLZJAKOV. Fundamentals of the Theory of Groups.
63 BOLLOBAS. Graph Theory.
64 EDWARDS. Fourier Series. Vol. I. 2nd ed.
65 WELLS. Differential Analysis on Complex Manifolds. 2nd ed.

66 WATERHOUSE. Introduction to Affine Group Schemes.
67 SERRE. Local Fields.
68 WEIDMANN. Linear Operators in Hilbert Spaces.
69 LANG. Cyclotomic Fields II.
70 MASSEY. Singular Homology Theory.
71 FARKAS/KRA. Riemann Surfaces. 2nd ed.
72 STILLWELL. Classical Topology and Combinatorial Group Theory. 2nd ed.
73 HUNGERFORD. Algebra.
74 DAVENPORT. Multiplicative Number Theory. 3rd ed.
75 HOCHSCHILD. Basic Theory of Algebraic Groups and Lie Algebras.
76 IITAKA. Algebraic Geometry.
77 HECKE. Lectures on the Theory of Algebraic Numbers.
78 BURRIS/SANKAPPANAVAR. A Course in Universal Algebra.
79 WALTERS. An Introduction to Ergodic Theory.
80 ROBINSON. A Course in the Theory of Groups. 2nd ed.
81 FORSTER. Lectures on Riemann Surfaces.
82 BOTT/TU. Differential Forms in Algebraic Topology.
83 WASHINGTON. Introduction to Cyclotomic Fields. 2nd ed.
84 IRELAND/ROSEN. A Classical Introduction to Modern Number Theory. 2nd ed.
85 EDWARDS. Fourier Series. Vol. II. 2nd ed.
86 VAN LINT. Introduction to Coding Theory. 2nd ed.
87 BROWN. Cohomology of Groups.
88 PIERCE. Associative Algebras.
89 LANG. Introduction to Algebraic and Abelian Functions. 2nd ed.
90 BRØNDSTED. An Introduction to Convex Polytopes.
91 BEARDON. On the Geometry of Discrete Groups.
92 DIESTEL. Sequences and Series in Banach Spaces.
93 DUBROVIN/FOMENKO/NOVIKOV. Modern Geometry—Methods and Applications. Part I. 2nd ed.
94 WARNER. Foundations of Differentiable Manifolds and Lie Groups.
95 SHIRYAEV. Probability. 2nd ed.
96 CONWAY. A Course in Functional Analysis. 2nd ed.
97 KOBLITZ. Introduction to Elliptic Curves and Modular Forms. 2nd ed.
98 BRÖCKER/TOM DIECK. Representations of Compact Lie Groups.
99 GROVE/BENSON Finite Reflection Groups. 2nd ed.
100 BERG/CHRISTENSEN/RESSEL. Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions.
101 EDWARDS. Galois Theory.
102 VARADARAJAN. Lie Groups, Lie Algebras and Their Representations.
103 LANG. Complex Analysis. 3rd ed.
104 DUBROVIN/FOMENKO/NOVIKOV. Modern Geometry—Methods and Applications. Part II.
105 LANG. $SL_2(\mathbf{R})$.
106 SILVERMAN. The Arithmetic of Elliptic Curves.
107 OLVER. Applications of Lie Groups to Differential Equations. 2nd ed.
108 RANGE. Holomorphic Functions and Integral Representations in Several Complex Variables.
109 LEHTO. Univalent Functions and Teichmüller Spaces.
110 LANG. Algebraic Number Theory.
111 HUSEMÖLLER. Elliptic Curves.
112 LANG. Elliptic Functions.
113 KARATZAS/SHREVE. Brownian Motion and Stochastic Calculus. 2nd ed.
114 KOBLITZ. A Course in Number Theory and Cryptography. 2nd ed.
115 BERGER/GOSTIAUX. Differential Geometry: Manifolds, Curves, and Surfaces.
116 KELLEY/SRINIVASAN. Measure and Integral. Vol. I.
117 SERRE. Algebraic Groups and Class Fields.
118 PEDERSEN. Analysis Now.
119 ROTMAN. An Introduction to Algebraic Topology.
120 ZIEMER. Weakly Differentiable Functions: Sobolev Spaces and Functions of Bounded Variation.
121 LANG. Cyclotomic Fields I and II. Combined 2nd ed.
122 REMMERT. Theory of Complex Functions. *Readings in Mathematics*
123 EBBINGHAUS/HERMES et al. Numbers. *Readings in Mathematics*
124 DUBROVIN/FOMENKO/NOVIKOV. Modern Geometry—Methods and Applications. Part III.
125 BERENSTEIN/GAY. Complex Variables: An Introduction.
126 BOREL. Linear Algebraic Groups. 2nd ed.
127 MASSEY. A Basic Course in Algebraic Topology.
128 RAUCH. Partial Differential Equations.
129 FULTON/HARRIS. Representation Theory: A First Course. *Readings in Mathematics*
130 DODSON/POSTON. Tensor Geometry.

131 LAM. A First Course in Noncommutative Rings. 2nd ed.
132 BEARDON. Iteration of Rational Functions.
133 HARRIS. Algebraic Geometry: A First Course.
134 ROMAN. Coding and Information Theory.
135 ROMAN. Advanced Linear Algebra.
136 ADKINS/WEINTRAUB. Algebra: An Approach via Module Theory.
137 AXLER/BOURDON/RAMEY. Harmonic Function Theory. 2nd ed.
138 COHEN. A Course in Computational Algebraic Number Theory.
139 BREDON. Topology and Geometry.
140 AUBIN. Optima and Equilibria. An Introduction to Nonlinear Analysis.
141 BECKER/WEISPFENNING/KREDEL. Gröbner Bases. A Computational Approach to Commutative Algebra.
142 LANG. Real and Functional Analysis. 3rd ed.
143 DOOB. Measure Theory.
144 DENNIS/FARB. Noncommutative Algebra.
145 VICK. Homology Theory. An Introduction to Algebraic Topology. 2nd ed.
146 BRIDGES. Computability: A Mathematical Sketchbook.
147 ROSENBERG. Algebraic K-Theory and Its Applications.
148 ROTMAN. An Introduction to the Theory of Groups. 4th ed.
149 RATCLIFFE. Foundations of Hyperbolic Manifolds.
150 EISENBUD. Commutative Algebra with a View Toward Algebraic Geometry.
151 SILVERMAN. Advanced Topics in the Arithmetic of Elliptic Curves.
152 ZIEGLER. Lectures on Polytopes.
153 FULTON. Algebraic Topology: A First Course.
154 BROWN/PEARCY. An Introduction to Analysis.
155 KASSEL. Quantum Groups.
156 KECHRIS. Classical Descriptive Set Theory.
157 MALLIAVIN. Integration and Probability.
158 ROMAN. Field Theory.
159 CONWAY. Functions of One Complex Variable II.
160 LANG. Differential and Riemannian Manifolds.
161 BORWEIN/ERDÉLYI. Polynomials and Polynomial Inequalities.
162 ALPERIN/BELL. Groups and Representations.
163 DIXON/MORTIMER. Permutation Groups.
164 NATHANSON. Additive Number Theory: The Classical Bases.
165 NATHANSON. Additive Number Theory: Inverse Problems and the Geometry of Sumsets.
166 SHARPE. Differential Geometry: Cartan's Generalization of Klein's Erlangen Program.
167 MORANDI. Field and Galois Theory.
168 EWALD. Combinatorial Convexity and Algebraic Geometry.
169 BHATIA. Matrix Analysis.
170 BREDON. Sheaf Theory. 2nd ed.
171 PETERSEN. Riemannian Geometry.
172 REMMERT. Classical Topics in Complex Function Theory.
173 DIESTEL. Graph Theory. 2nd ed.
174 BRIDGES. Foundations of Real and Abstract Analysis.
175 LICKORISH. An Introduction to Knot Theory.
176 LEE. Riemannian Manifolds.
177 NEWMAN. Analytic Number Theory.
178 CLARKE/LEDYAEV/STERN/WOLENSKI. Nonsmooth Analysis and Control Theory.
179 DOUGLAS. Banach Algebra Techniques in Operator Theory. 2nd ed.
180 SRIVASTAVA. A Course on Borel Sets.
181 KRESS. Numerical Analysis.
182 WALTER. Ordinary Differential Equations.
183 MEGGINSON. An Introduction to Banach Space Theory.
184 BOLLOBAS. Modern Graph Theory.
185 COX/LITTLE/O'SHEA. Using Algebraic Geometry.
186 RAMAKRISHNAN/VALENZA. Fourier Analysis on Number Fields.
187 HARRIS/MORRISON. Moduli of Curves.
188 GOLDBLATT. Lectures on the Hyperreals: An Introduction to Nonstandard Analysis.
189 LAM. Lectures on Modules and Rings.
190 ESMONDE/MURTY. Problems in Algebraic Number Theory.
191 LANG. Fundamentals of Differential Geometry.
192 HIRSCH/LACOMBE. Elements of Functional Analysis.
193 COHEN. Advanced Topics in Computational Number Theory.
194 ENGEL/NAGEL. One-Parameter Semigroups for Linear Evolution Equations.
195 NATHANSON. Elementary Methods in Number Theory.
196 OSBORNE. Basic Homological Algebra.
197 EISENBUD/HARRIS. The Geometry of Schemes

198 ROBERT. A Course in p-adic Analysis.
199 HEDENMALM/KORENBLUM/ZHU. Theory of Bergman Spaces.
200 BAO/CHERN/SHEN. An Introduction to Riemann–Finsler Geometry.
201 HINDRY/SILVERMAN. Diophantine Geometry: An Introduction.
202 LEE. Introduction to Topological Manifolds.
203 SAGAN. The Symmetric Group: Representations, Combinatorial Algorithms, and Symmetric Functions.
204 ESCOFIER. Galois Theory.
205 FÉLIX/HALPERIN/THOMAS. Rational Homotopy Theory. 2nd ed.
206 MURTY. Problems in Analytic Number Theory.
Readings in Mathematics
207 GODSIL/ROYLE. Algebraic Graph Theory.
208 CHENEY. Analysis for Applied Mathematics.
209 ARVESON. A Short Course on Spectral Theory.
210 ROSEN. Number Theory in Function Fields
211 LANG. Algebra, 3rd ed.

Serge Lang

Algebra

Revised Third Edition

This book is intended as a basic text for a one-year course in algebra at the graduate level, or as a useful reference for mathematicians and professionals who use higher-level algebra. It successfully addresses the basic concepts of algebra. For the revised third edition, the author has included additional exercises and made numerous corrections to the text.

Comments on Serge Lang's *Algebra*:

"Lang's *Algebra* changed the way graduate algebra is taught, retaining classical topics but introducing language and ways of thinking from category theory and homological algebra. It has affected all subsequent graduate-level algebra books."

—*From April 1999* Notices of the AMS, *announcing that the author was awarded the Leroy P. Steele Prize for Mathematical Exposition for his many mathematics books*

"The author has an impressive knack for presenting the important and interesting ideas of algebra in just the "right" way, and he never gets bogged down in the dry formalism which pervades some parts of algebra."

—*From* MathSciNet's *review of the first edition*

ISBN 0-387-95385-X

springeronline.com